# Kielentutkimuksen menetelmiä

# I–IV

*Toimittaneet*

MILLA LUODONPÄÄ-MANNI
MARKUS HAMUNEN
REETTA KONSTENIUS
MATTI MIESTAMO
URPO NIKANNE
KAIUS SINNEMÄKI

SKS

# Kielentutkimuksen menetelmiä I–V

*Toimittaneet*

MILLA LUODONPÄÄ-MANNI, MARKUS HAMUNEN,
REETTA KONSTENIUS, MATTI MIESTAMO,
URPO NIKANNE JA KAIUS SINNEMÄKI

SUOMALAISEN KIRJALLISUUDEN SEURA • HELSINKI • 2020

SUOMALAISEN KIRJALLISUUDEN
SEURAN TOIMITUKSIA 1457

Teos on Suomalaisen Kirjallisuuden Seuran
nimeämien asiantuntijoiden tarkastama.

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus



Kannen suunnittelu: Timo Numminen
Taitto ja EPUB: Keski-Suomen Sivu Oy

ISBN 978-951-858-173-7 (I–IV, nid.)
ISBN 978-951-858-180-5 (PDF)
ISBN 978-951-858-178-2 (EPUB)

ISSN 0355-1768 (nid.)
ISSN 2670-2401 (verkkojulkaisut)

DOI: https://doi.org/10.21435/skst.1457

# Alkusanat

*Kielentutkimuksen menetelmiä* on useita aloja käsittävä kielentutkimuksen menetelmiin keskittyvä teos opiskelijan ja tutkijan tueksi suomen kielellä. Teoksen keskeisenä päämääränä on esitellä ja selkiyttää menetelmällisten eli metodisten mahdollisuuksien kirjoa, menetelmällisten valintojen syitä ja seurauksia, menetelmien vahvuuksia ja heikkouksia sekä ylipäätään metodologisen ja laajemmin metateoreettisen tausta-ajattelun ilmaisemisen tärkeyttä kielentutkimuksessa.

Perus- ja jatko-opiskelijoilta saadun palautteen perusteella kokoelma tulee tarpeeseen, sillä tieteenfilosofinen, metodologinen ja kielentutkimuksen menetelmiä koskeva yleinen teoreettinen koulutus on Suomessa paikoin turhan ohutta ja pirstoutunut eri suuntauksiin ja oppiaineisiin. Tämä eri yliopistoja, oppiaineita ja suuntauksia edustavien tutkijoiden yhteistyössä laatima hanke on yksi keino torjua kielentutkimuksen menetelmien ja metateorian opetuksen pirstaloitumista.

Vaikka vakiintuneen normaalitieteen (ks. Kuhn 1962) harjoittaminen on tieteellisen tiedon luomiseksi välttämätöntä, tutkijoiden on syytä perehtyä myös metodologiaan ja metateoriaan, jotta tiede uudistuisi ja kehittyisi. Metodologia harjaannuttaa eksplisiittiseen metodien kriittiseen arviointiin ja perusteluun, ja metateoria selvittelee tutkimuksen taustoja, logiikkaa, rajauksia ja historiaa sekä muita tutkimukseen usein piilevästi vaikuttavia tekijöitä. Näitä asioita ei voi jättää hiljaisen tiedon varaan. Kielitieteen menetelmien ja metateorian laajaa esittelyä tarvitaan

myös välttämään suomalaisen kielentutkimuksen eklektisyyttä eli erilaisten tutkimusperinteiden ja teorioiden yhdistelyä tutkimustyössä. Tällaisen yhdistelyn vuoksi tutkijan on kyettävä arvioimaan käyttämiensä menetelmien taustateorioita ja niiden keskinäistä yhteensopivuutta. Eklektisyys ei ole kuitenkaan sama asia kuin monimenetelmäisyys tai monimenetelmäinen tutkimusasetelma, jotka kumpikin perustuvat tietoisiin, ilmipantuihin metateoreettisiin ratkaisuihin. Eklektisyyttä voi kuvata teorioiden ja mallien käyttämisenä yhdessä ja samassa tutkimuksessa välttämättä tietämättä, ovatko ne keskenään ristiriidattomia.

Kielentutkimuksen metateoria tarkastelee siis kielentutkimusta nimensä mukaisesti metatasolta. Metateorian voi jakaa normatiiviseen ja faktuaaliseen tai deskriptiiviseen tutkimukseen. Normatiivista tutkimusta on tieteenfilosofia, johon sisältyvät metodologia ja tieteen etiikka. Se pohtii sitä, millaisia piirteitä tieteellisessä tutkimuksessa tulee olla, jotta se täyttäisi tieteellisyyden kriteerit. Faktuaalista tai deskriptiivistä metateoriaa taas on tieteentutkimus, esimerkiksi kielitieteen sosiologia, psykologia ja historia. Deskriptiivistä metateoriaa ei teoksessa juurikaan esitellä, sillä tämän kokoelman tavoite on nimenomaan tarjota lukijalle käsitteellisiä välineitä kielentutkimuksen tutkimusmetodien ja taustateorioiden arviointiin. Kielitieteen historiaa käsitellään artikkeleissa jonkin verran esiteltävän menetelmän ymmärtämiseksi. Kielentutkimuksen sosiologia ja psykologia jäävät käytännössä teoksen rajauksen ulkopuolelle.

Yleisesti etuliitteellä *meta-* kuvataan ylempää tai abstraktimpaa näkökulmaa johonkin tarkasteltavaan asiaan. Esimerkiksi *metatieto* tarkoittaa tietoa tiedosta (esim. tutkimusaineiston metatiedot). *Metakieli* tarkoittaa toisen asteen kieltä, jolla puhutaan luonnollisesta kielestä tai objektikielestä. Esimerkiksi perinteinen kielenkuvausten kielioppitermistö on metakieltä usein jostakin yksittäisestä objektikielestä. *Metateoria* tarkoittaa vanhastaan teoriaa teorioista. *Metatutkimus* on tutkimus, joka perustuu joukkoon muita, jotakin spesifiä aihetta käsitteleviä tutkimuksia, ja niin edelleen. Pidämme tärkeänä, että tarjolla on sellainen kielentutkimuksen menetelmällistä kirjoa tarkasteleva suomenkielinen teos, joka juuri metatasoisen käsittelyn kautta pyrkii tarjoamaan välineitä metodologisten ongelmien arviointiin.

Kielentutkimusta harjoitetaan erilaisten suuntausten piirissä ja niissä erilaisin eriytynein menetelmin. Kielentutkimuksen suuntaukset ja niiden parissa käytettävät tutkimusmenetelmät ovat kuitenkin siinä määrin kytköksissä toisiinsa, että niitä on usein lähes mahdotonta esitellä erikseen. Tähän vaikeuteen törmäsimme useaan otteeseen kokoelmaa valmistaessamme. Olemmekin joutuneet mahdollisesti tekemään tieteenalan totutulle jäsennykselle jonkinasteista väkivaltaa, sillä olemme pyrkineet nimenomaan tarkempaan kielentutkimuksen metodologian kuvaukseen. Tämä pyrkimys on tarkoittanut sitä, että keinotekoisia rajoja tutkimusmenetelmien ja suuntausten välillä on tehty tai jätetty tekemättä silloin, kun jotakin relevanttia metodologista metatietoa on kirjoittajien mielestä näin saatu tuotua esiin.

Pyrkiessämme kuvaamaan kielentutkimuksen menetelmällistä kirjoa olemme yksittäisten menetelmien (kuten luonnollisen kielenkäytön havainnointi tai kielitaju) lisäksi ottaneet mukaan myös joitain yksittäisten tutkimussuuntausten (kuten sosiolingvistiikka tai kielitypologia) kuvauksia silloin, kun näiden käyttämiä menetelmiä on mahdotonta käsitellä ilman tutkimussuuntauksen tuntemusta. *Kielentutkimuksen menetelmiä* ei siis tässä mielessä ole sama asia kuin esimerkiksi ”Kielentutkimuksen käsikirja” tai ”Kielentutkimuksen perusteet”, emmekä ole tässä teoksessa pyrkineet kuvaamaan kattavasti koko kielentutkimuksen tutkimussuuntausten ja tutkimusalojen kirjoa. Olemme hyvin tietoisia, että monia kielentutkimukselle keskeisiä aloja jää pääasiassa käytännön syistä tämän kokoelman ulkopuolelle.

Ensiksi kokoelman tavoitteena ei ole tarjota yleiskuvaa kielitieteen sellaisista perustavista osa-alueista kuin fonetiikka, fonologia, morfologia, syntaksi, leksikologia, semantiikka tai pragmatiikka. Näitä esitellään kielentutkimuksen perusoppikirjoissa (esim. Karlsson 1998; Ojutkangas, Larjavaara, Miestamo & Ylikoski 2009) ja harjoitutetaan käytännön taidoksi kielitieteen peruskoulutuksessa. Näin ollen sellaisia analyysitekniikoita ja -perinteitä kuin foneemianalyysi (esim. yksiköiden jakauma ja substituutio), morfeemianalyysi (esim. morfemisointi), syntaksin operationaaliset testit (esim. siirtotesti) sekä sanasemanttinen analyysi (homonymia, synonymia, polysemia), sanaluokkaluokittelu, semanttinen piirreanalyysi yms. ei käsitellä kattavasti tässä teoksessa.

Tulee kuitenkin muistaa, että tällainen kielen perusanalyysikin on pohjimmiltaan teoreettis-metodista toimintaa. Siksi teoksen tarjoamaa metateoreettisesti ja metodologisesti orientoitunutta lähestymistapaa kannattaa lukijan soveltaa myös tällaiseen perustutkimukseen.

Toiseksi edellä mainituista kielitieteen perustavista osa-alueista on olemassa jo erittäin runsas alakohtainen peruskirjallisuus. Esimerkiksi fonologian ja morfologian kuvausmalleja on lukuisia, syntaksin kymmeniä. Lisäksi erilliset metodioppaat (esim. Tainio 1997) tai muut alakohtaiset yleisesitykset ovat usein tarkoituksenmukaisempia paikkoja metodisten yksityiskohtien käsittelyyn kuin valitsemamme yleisemmällä tasolla liikkuva käsikirjamainen esitystapa. Tämän kokoelman tavoitteena on antaa yleiskatsaus kielentutkimuksen metodioppiin ja sen peruslähteisiin, ja siksi teoksen luvut pyrkivät haravoimaan kielentutkimuksen kenttää pragmaattisesti metodologisesta näkökulmasta. Kieliaineistot, kuvaustekniikat (tai metodit) ja kuvauskäsitteet muodostavat kielitieteen kaikessa toiminnassa läpikäyvän kolmikannan, jonka perusteita tieteenalan metodologia erittelee. Tämä kokoelma tarjoaa tähän yleiskuvan. Yksityiskohdissa lukijan on hyvä kääntyä erikoisalan kirjallisuuden puoleen (ks. Johdannon luku 5 tk.).

Tämän kokoelman sisältö on joka tapauksessa pyritty valitsemaan pragmaattisin perustein, ilman arvottavia latauksia. Samalla tavalla pidämme arvottamista tarpeettomana myös teoksessa kuvattujen eri menetelmien välillä. Sen sijaan pidämme kielentutkimuksen menetelmiä toisiaan täydentävinä ja pyrimme kuvaamaan niitä niiden omista lähtökohdista käsin ja tunnistamaan niiden sovellusalat, heikkoudet ja vahvuudet. On huomattava, että erilaisilla metodeilla tavoitellaan erilaista tietoa. Metodeilla on siten erilaisia heikkouksia ja vahvuuksia ainoastaan ja erityisesti suhteessa asetettuihin tutkimuskysymyksiin. Tämän johdantoteoksen tavoitteena on tuoda opiskelijan ja tutkijan ulottuville tällaista olemassa olevaa standarditietoa metodeista.

Terminologisesti kokoelman tavoitteena on objektiivinen, läpinäkyvä ja ei-arvottava esitys nykymetodologiassa käytetyistä käsitteistä. Pyrimme painottamaan metodologian ja tieteenfilosofian nykykäsitteistöä. Tältä osin erityisesti kokoelman verraten laajan johdannon tavoitteena on tarjota myös uusia virikkeitä, koska tieteenfilosofiasta kirjoitetaan

suomalaisessa ja kansainvälisessä kielentutkimuksessa suhteellisen vähän. Metodologiaa koskeva termistö on määritelty keskeisiä tieteellisen käsitteenmuodostuksen perusperiaatteita noudattaen ja kattamaan koko teoksen (ks. Johdannon luku 1 tk.). Kokoelman luvuissa eri kirjoittajat ilmaisevat erikseen, mikäli heidän metodinsa puitteissa käytetään termejä jossakin eriävässä merkityksessä.

Kokoelmassa omaksuttu metateoreettinen lähestymistapa kulkee punaisena lankana läpi teoksen kaikkien lukujen. Teoksen jokaisessa luvussa pohditaan kyseisen menetelmän tai tutkimussuuntauksen esittelyn yhteydessä sen historiaa, suhdetta lähialoihin sekä mahdollisia erityisiä eettisiä kysymyksiä. Tällaisen pohdinnan tarkoituksena on auttaa lukijaa hahmottamaan paremmin kielentutkimuksen ja sen menetelmien kokonaisuutta. Metateoreettinen lähestymistapa näkyy lisäksi kokoelman johdanto-osion kattavuutena ja runsaina teoksen sisäisinä viittauksina. Johdannossa korostetaan ja tuodaan konkreettisesti esiin myös metodologisen tason kysymysten tärkeyttä itsessään tavanomaisten ilmiötutkimusten rinnalla.

Kokoelma on suunnattu kielten perus- ja jatko-opiskelijoille, tutkijoille ja muille kielentutkimuksesta kiinnostuneille. Se valaisee kielentutkimuksessa käytettäviä menetelmiä ja niihin liittyviä tutkimusprosesseja mahdollisimman käytännöllisesti ja havainnollisesti sekä avaa monia menetelmäkirjallisuudessa aikaisemmin vähemmälle huomiolle jääneitä kohtia. Teoksessa käsitellään monia ajankohtaisia tieteen yhteiskunnalliseen asemaan liittyviä aiheita, kuten tutkimusetiikkaan, tietosuojaan ja avoimen tieteen periaatteisiin liittyviä kysymyksiä. Erityisen hyvin teos soveltuu kielentutkimuksen menetelmäopetukseen, mutta se esittelee myös laajempia tieteelliseen tutkimukseen yleisesti liittyviä tieteenfilosofisia taustoja.

Toimituskuntaa on johtanut Milla Luodonpää-Manni Markus Hamusen avustamana.

Kokoelmaa laadittaessa osia siitä on lukenut ja kommentoinut koko joukko asiantuntijoita. Haluamme kiittää seuraavia ihmisiä – panoksesta kommentoivana esilukijana tai muusta hyödyksi koituneesta neuvonannosta – sukunimen mukaisessa aakkosjärjestyksessä: Lotta Aarikka, Sonja Dahlgren, Anton Granvik, Viljami Haakana, Auli

Hakulinen, Sointu Järvinen, Annekatrin Kaivapalu, Fred Karlsson, Markus Lakaniemi, Jorma Luutonen, Arto Mustajoki, Urho Määttä, Tiina Onikki-Rantajääskö, Jarno Porkka, Jesse Salmi, Antonio Sciacovelli, Tuulia Sinervo, Ulla Vanhatalo ja Camilla Wide. Lisäksi kiitämme Viljami Haakanaa lähdeluetteloiden tarkistamisesta sekä Nora Fabritiusta kaavioiden tekemisestä lukuun 15. Erikseen toimitus kiittää koko kirjoittajakuntaa vakuuttavasta sitoutumisesta laajaan yhteiseen hankkeeseen. Lämmin kiitos myös kustantajalle. Teoksen valmistumista on tukenut Otto A. Malmin lahjoitusrahasto sekä Helsingin yliopiston Nykykielten laitos (nykyinen Kielten osasto).

Hyviä lukuhetkiä toivottaen

Toimittajat

# Sisällys

# Osa I
# Johdanto

# 1. Tutkimuksen käytäntö

*Milla Luodonpää-Manni*
https://orcid.org/0000-0002-1646-4462
*Markus Hamunen*
https://orcid.org/0000-0003-2786-6392
*Reetta Konstenius*

Tässä luvussa keskitymme tutkimuksenteon käytäntöön.* Etenemme tutkimuksenteon taustalla vaikuttavien tieteenfilosofisten ja metodologisten periaatteiden esittelystä aiheen valintaan, tutkimusstrategiaan, aineistonkeruuseen, aineiston analyysiin ja tulosten esittelyyn liittyviin kysymyksiin. Lopuksi pohdimme vielä, miten kielitieteen menetelmillä voidaan vastata ihmiskunnan suuriin haasteisiin. Johdannon luvuissa 2 ja 3 käsittelemme syvällisemmin tieteellisen menetelmän periaatteita ensin yleisellä tasolla ja sitten erityisesti kielentutkimuksen näkökulmasta. Kielentutkimuksen etiikkaan syvennymme luvussa 4, joskin etiikkaan liittyviä aiheita sivutaan lyhyesti myös luvuissa 1 ja 2. Päätämme johdannon lukuun 5, jossa esittelemme tämän teoksen rakennetta sekä muita kielentutkimuksen metodioppaita.

* Kiitämme ulkopuolisia arvioijia ja kommentoijia sekä kokoelman muuta toimituskuntaa koko johdantoa koskevista hyödyllisistä kommenteista.

## 1.1. Kielentutkimuksen kolme tasoa

Tutkijan omaksumilla tieteenfilosofisilla ja metodologisilla periaatteilla on suuri vaikutus tutkimuksenteon käytäntöön, esimerkiksi siihen, miten aiheita, tutkimuskohteita ja menetelmiä valitaan ja käytetään. *Tieteenfilosofiaksi* kutsutaan sitä periaatteiden ja käsitysten joukkoa, joka ohjaa tieteen tekemistä. Tieteenfilosofia on luonteeltaan normatiivista ja kertoo, miten tiedettä tulisi tehdä. Se eroaa tieteentutkimuksesta, joka on deskriptiivistä ja joka kuvailee tiedeyhteisössä tapahtuvaa tosiasiallista toimintaa. Vastaava ero voidaan tehdä myös yksittäisen tieteenalan sisällä. Kielitieteen filosofia siis pohtii kielitieteen tutkimuskohteiden valintaa ja kielitieteessä käytettäviä menetelmiä tieteenfilosofian näkökulmasta ja kielentutkimuksen tutkimus, kuten kielentutkimuksen sosiologia tai historia, mitä kielentutkimuksessa tosiasiassa tehdään tai on tehty.

On korostettava, että esimerkiksi kielentutkimuksen filosofia määrittelee kielentutkimuksen yleiset pelisäännöt ja ne kriteerit, joilla arvioidaan tieteenalan menetelmien tieteellisyyttä. Kielentutkimuksen tutkimuksessa voidaan arvioida, kuinka näitä menettelytapoja noudatetaan. Kumpikaan näistä metateoreettisista tutkimusaloista ei säätele esimerkiksi tutkimuksen yksittäisiä aihevalintoja tai tutkimuskysymyksiä, vaan niissä on kiinnostuttu muun muassa siitä, kuinka muotoillaan tutkimuskysymyksiä, kuinka tutkitaan tieteellisesti ja kuinka tutkimuksen tuloksia tulkitaan. Edellinen tarkastelee tätä normin, jälkimmäinen tosiasiallisen toiminnan kannalta.

Sekä kielentutkimuksen filosofialla että kielentutkimuksen tutkimuksella on metanäkökulma kielentutkimukseen, ja tutkimusalueita voidaan nimittää yhteisellä nimellä *kielentutkimuksen metateoria*. Alla olevassa kuviossa 1 kieli ja sen tutkimus jaetaan neljään tasoon. Tason 1 muodostaa kielen ilmitodellisuus, josta raakadata tutkimukseen kerätään. Itse kielentutkimus operoi tasoilla 2–4. Metateorian tasolta 4 kuvataan ja selitetään kielentutkimuksen teorioita ja metodeja (3). Tasolla 3 kuvatut teoriat ja metodit puolestaan koskevat kahden alemman tason muodostamaa kokonaisuutta eli kieltä sellaisena kuin se reaalitodellisuudessa ilmenee (1) kielentutkimuksesta riippumatta sekä

kielenilmiöitä (2), jotka kielentutkimuksen yhteisö tai yksittäinen tutkija tunnistaa ja käsitteellistää tieteellisen kuvauksen, selitysten tai teorioiden kohteeksi.

Metateorian tutkimuskysymyksen voi kiteyttää seuraavasti: mitä on kielentutkimus, joka tarkastelee sitä, mitä on kieli? Metateoreettisella tutkimuksella on kielitieteessä usein juuri tarkoitettu esimerkiksi kahden teoreettisen viitekehyksen, kuten kognitiivisen kielentutkimuksen ja generatiivisen kielentutkimuksen, perusteiden keskinäistä arviointia ja vertailua tai jonkin yksittäisen tutkimusalan perusteiden selvittelyä (jälkimmäisestä ks. esim. Määttä 1994).

Metateoria tulee erottaa kielentutkimuksen lähestymistavoista, koulukunnista ja paradigmoista, joissa myös toteutetaan tietynlaista tutkimusperspektiiviä ja rajattua näkökulmaa kieleen ja tämän myötä usein myös valikoitunutta metodiikkaa. Koulukunnat rakentuvat usein ajallis-paikallisesti keskeisten tutkijoiden ja heidän seuraajiensa ympärille. Ne nimetäänkin usein prominentin tutkijan tai paikan mukaan (ks. esim. Sampson 1980). Termi *paradigma* on lähtöisin tieteenfilosofi Thomas Kuhnilta (ks. Kuhn 1994 [1962]; ks. myös Kiikeri & Ylikoski 2004, 55–69). Hän tarkoittaa sillä tieteenalalla vakiintunutta ja jaettua näkemystä eli tieteellistä konsensusta siitä, mitä pidetään tutkimisen arvoisena, millaisia kysymyksiä esitetään ja miten tutkimusongelmia ratkaistaan ja tuloksia esitetään ja tulkitaan (myös Percival 1976). Tällaisiin valintoihin ja näkemyksiin liittyvät myös erilaiset tieteen tekemisen ohjausvaikutukset, tieteellinen vallankäyttö ja muut subjektiivisetkin tekijät (esim. Karlsson 1975).

Metodologia eli menetelmäoppi on osa tieteenfilosofiaa, ja sillä tarkoitetaan tieteellisen metodin määrittelemistä, kehittämistä ja soveltamista koskevien periaatteiden ja ehtojen tutkimusta. Metodologiassa siis tutkitaan tieteellistä menetelmää eli keinoja, joilla tieteellistä tietoa hankitaan. Se on osa metateoriaa ja sijoittuu kuviossa 1 tasolle 4. Metodiikka taas on jollakin tieteenalalla käytössä olevien menetelmien, tutkimustekniikoiden ja välineiden joukko. Siihen kuuluvat ainakin aineiston keräämisen, kuvaamisen ja analyysin sekä päätelmien tekemisen tekniikat.

*Metodeilla* eli tutkimusmenetelmillä viitataan niihin erityisiin keinoihin, välineisiin ja tutkimustekniikkoihin, joiden avulla tutkimus-

METATEORIA (4)

NORMATIIVINEN: kielentutkimuksen tieteenfilosofia, joka sisältää metodologian ja tutkimusetiikan

FAKTUAALINEN: kielentutkimuksen historia, sosiologia, psykologia yms.

↓↑

TEORIAT JA METODIT (3)

esim. generatiivinen kielioppi, historiallis-vertaileva menetelmä, vuorovaikutuslingvistiikka, lingvistinen kenttätyöskentely, systeemis-funktionaalinen kielioppi, kielitypologia, haastattelu, laadullinen tutkimus, kognitiivinen kielioppi, tilastomenetelmät, kielitaju, naturalistinen observointi, maallikkokielenpuhujan konsultointi (eli elisitaatio), aineiston litterointitavat, muut ilmiön ja tulosten kuvaamistavat (taulukot, kuviot, formalisointi ja notaatiotavat yms.)jne.

↓↑

KIELENILMIÖT (2)

esim. ranskan subjunktiivin käyttö, suomen -*kin*-liitepartikkeli puheessa, duaali maailman kielissä, ajan liikemetaforat slaavilaiskielissä, pörssitiedotteiden kielenkäyttö, Juvan vanhimmat paikannimet, futuurisuuden ilmaiseminen Ruotsin viittomakielessä, selkokielen lauserakenteet, nykytietokirjallisuuden kääntäminen, silmien liike luettaessa jne.

↓↑

KIELI (1)

puhuminen ja kuunteleminen tai viittominen ja viittomien tulkinta (kieli toimintana, transitorisena prosessina)
kirjoittaminen ja lukeminen (kieli toimintana, joka tuottaa tekstin)
puhutut ja kirjoitetut tekstit (kieli objektina, reifikaatio)
kieltajun ilmentymät

Kuvio 1. Kieli ilmentymineen (1) sekä kielentutkimuksen kolme tasoa (2–4).

kohteesta pyritään hankkimaan tieteellistä tietoa. Kielentutkimuksessa menetelmällisiä välineitä tarjoavat myös monet teoriat ja mallit (ks. kuvion 1 tasoa 3). Metodien valintaan vaikuttavat ainakin tutkimuskohde ja tutkimuskysymykset sekä tutkimuksen taustalla olevat metodologiset periaatteet. Yksittäisistä metodeista ei siten ole mielekästä puhua pohtimatta tutkimuksenteon taustalla vaikuttavaa laajempaa menetelmäoppia. Olemmekin pyrkineet avaamaan keskeisimpiä tieteenfilosofisia ja metodologisia periaatteita johdannon luvussa 2.

On huomattava, että käytännössä metodologian ja metodiikan erossa pitäminen voi olla hankalaa ainakin kahdesta syystä. Ensinnäkin tieteenalan eri lohkojen taustateoriat ovat saattaneet rakentua siinä määrin eri lähtökohdista, että termejä on voitu käyttää hieman eri merkityksissä eri suuntausaloilla. Toiseksi metodologiaa ei usein voi käsitellä ilman metodiikkaa ja toisin päin. Näistä syistä teoksen lukujen kesken on

mahdollista huomata terminologisia konnotaatio- ja sävyeroja, vaikka tavoitteenamme on noudattaa tässä annettuja standardimääritelmiä.

Kuviossa 1 kieli (1) ja kielenilmiöt (2) on erotettu toisistaan. Kieli on monitahoinen ilmiö, jonka avulla ihminen ajattelee, viestii ja toimii vuorovaikutuksessa ympäröivän maailman kanssa. Kieli (1) ilmenee toimintana, kuten puhumisena ja viittomisena, kuuntelemisena ja viittomien havaitsemisena ja tulkitsemisena sekä kirjoittamisena ja lukemisena. Kieleen kuuluvat myös puhutut ja kirjoitetut tekstit. Tällä tasolla toimitaan kielellisesti arkisessa elämässä: vuorovaikutuksessa toisten kanssa, tekstiä luettaessa, puhetta suunniteltaessa, sanavalintoja mietittäessä jne. Tämä taso 1 on olemassa todellisuudessa kielentutkimuksesta huolimatta, ja sen voidaan ajatella muodostavan todellisuuden tason, johon kielentutkimus viime kädessä kohdistuu. Tasojen 1 ja 2 tosiasiallisessa erossa pitämisessä on samoja hankaluuksia kuin metodologian ja metodien kesken edellä, mutta käsitteellisenä erontekona se on tarpeen.

Kun kieli (1) jossain edellä kuvatussa ilmenemismuodossaan valitaan tutkittavaksi, siitä tulee kielenilmiöiden (2) tutkimusdataa. Tavanomaisen kieliopin perusrakennetta kuvaavan kielentutkijan (engl. *ordinary working grammarian*) teoreettis-metodinen toiminta esimerkiksi naturalistisessa havainnoinnissa perustuu kieli-ilmiöiden tunnistamiseen, luokittamiseen ja problematisoimiseen. Kielenilmiöt teoretisoituvat, abstrahoituvat ja pelkistyvät, sillä alkuperäinen kielenaines (raakadata, ks. Johdannon luku 1.4. tässä kirjassa) saatetaan tutkimusaineistoksi erilaisin metodisin ja käsitteellisin keinoin. Tämä koskee erityisesti puhuttua kieltä. Kun ohimenevää ja kertaluonteista kielenkäyttötilannetta ääni- tai videotallennetaan myöhemmin tarkasteltavaksi, alkuperäisestä tilanteesta menetetään aina jotain. Videokuvaus tapahtuu jollakin kuvanrajauksella sekä jostakin näkökulmasta. Äänitallenteessa kuullaan vain äänet. Lisäksi erilaiset tallennus- ja informaationpakkausmuodot sekä aineiston siirto- ja konvertointityöt saattavat redusoida dataa lisää. Esimerkiksi käytetty tiedonpakkausmenetelmä voi heikentää tallenteen äänenlaatua ja vaikuttaa siihen, millaista tutkimusta aineiston varassa voidaan tehdä (ks. Kokeellinen fonetiikka tk.). Analysoinnin helpottamiseksi puhutun kielen aineistoja, eli ääni- ja

videodataa, yleensä myös tekstinnetään (litteroidaan, transkriboidaan, tarkekirjoitetaan) luettavaan muotoon jollakin enemmän tai vähemmän vakiintuneella merkintätavalla. Kielentutkimuksen eri aloilla tähän on erilaisia transkriptio- ja litterointimenetelmiä ja -perinteitä, joissa esimerkiksi vuorovaikutuksen eri puolia (päällekkäispuhunta, tauot, eleet, ilmeet, asennot yms.) tai puheen foneettisen laadun tarkkuusasteita (erilaiset foneettiset tarkkeet) voidaan merkitä systemaattisesti näkyviin (ks. Haastattelu aineistonkeruun metodina sosiolingvistiikassa; Keskustelunanalyysi ja vuorovaikutuslingvistiikka tk.). Litterointitapa ja transkription hienous tai karkeus määräytyvät usein tutkimuskysymyksistä tai jonkin laajemman korpuksen koontiperiaatteista. Puheen tekstintäminen kuitenkin joka tapauksessa muuttaa, abstrahoi ja pelkistää aineistoa merkittävästi sen muuttaessa olomuotoaan puheesta tekstiksi. Matka kielen tasolta 1 (esim. puheen virrasta) teorioiden ja metodien tasolle 3 johtaa aina kielenilmiöiden (2) idealisoitumiseen sekä katsantotavan abstrahoitumiseen ja teoretisoitumiseen.

Edellä kuvatut tekstintämismetodit koskevat usein äidinkieltä tai tilannetta, jossa kielen perusrakenne tunnetaan jo. Toinen, edellistä perustavampi vaihe tekstintämisessä on se, kun aikaisemmin kuvaamattoman kielen kieliopin perusrakennetta kuvataan ensimmäistä kertaa. Tämä käsittää ennen kaikkea yksittäiskielen kaksoisjäsennyksen avaamista (ks. Martinet 1964, 22–24). Kaksoisjäsennys on luonnollisen kielen ominaisuus. Ensimmäinen jäsennys koskee sitä, että puheen virrasta tunnistetaan ja erotellaan kielen muoto-merkitysyksiköt eli morfeemit, joista laaditaan luettelo ja keskinäistä yhdistämistä koskevat säännöt (metodista esim. Nida 1949; Harris 1957 [1951], luvut 12–19). Toinen jäsennys koskee muoto-merkitysyksiköiden rakenneanalyysia pienempiin yksiköihin eli merkitystä erotteleviin foneemeihin, kielen foneemijärjestelmän hahmottamista ja foneemien keskinäistä yhdistämistä koskevien sääntöjen laatimista (metodista esim. Pike 1966 [1947]). Kaksoisjäsennyksen analyysi on jo kielen teoreettista kuvaamista. Se on myös edellytys puheen litteroimiselle tai transkriboimiselle esimerkiksi kielellisen vuorovaikutuksen ja sosiolingvististen seikkojen tutkimiseksi.[1] (Ks. Johdatus kielitieteeseen, Ojutkangas, Larjavaara, Miestamo & Ylikoski 2009.)

Kuviossa 1 kielenilmiöiden tasoa 2 voidaan kutsua siksi yleistettävien ilmiöiden joukoksi, jonka kielentutkimuksen yhteisö tunnistaa tieteellisen kuvauksen, selitysten tai teorioiden kohteeksi. Yksittäisten kielenilmiöiden tasolle tultaessa kielen teoreettis-metodista analyysia on siis jo tutkimuksen alkamishetkellä välttämättä olemassa. Esimerkiksi ranskan subjunktiivin käytön tutkiminen edellyttää monien teoreettisten termien ja väitteiden omaksumista. Väitteessä ”ranskan verbi *veuille* on subjunktiivi” muoto *veuille* on kaksoisjäsennyksen analyysin tulos ja *ranska* (kieli), *verbi* sekä *subjunktiivi* ovat termejä teoreettisille käsitteille, jotka on erikseen määriteltävä. Tällaiset yksittäiset väitteet ovat oikeastaan jo minimaalisia teorioita (teoriapartikulaareja), joiden muodostamista kokonaisuuksista muodostuu kielten teoreettisia osa- tai kokonaiskuvauksia (ks. Kielten dokumentointi ja kieliopin kuvaus; Kieliopin tutkimus ja kielioppiteoriat tk.). Kielentutkimuksen kuvausten, selitysten ja teorioiden kohteena olevat kielenilmiöt (2) eroavat siten arkisessa elämässä esimerkiksi puheen virtana havaittavasta, lingvistisesti analysoimattomasta kielenaineksesta (1).

**Tässä teoksessa käytettyjen metodologisten termien merkitykset**

**Meta**- on etuliite, jolla kuvataan ylempää abstraktimpaa näkökulmaa kohteeseen. Metatieto tarkoittaa tietoja tiedosta, metakieli toisen asteen kieltä (K2), jolla puhutaan luonnollisesta kielestä (K1), metateoria tarkoittaa teoriaa teorioista, metamatematiikka tutkii matematiikan alan kysymyksiä jne.

**Tieteenfilosofia** tutkii tieteen ja tieteellisen tiedon ehtoja. Tieteenfilosofia on normatiivista, ja se kertoo, miten tiedettä tulisi tehdä. Yleisen tieteenfilosofian lisäksi siitä voidaan erottaa yksittäisiä tieteenaloja koskevia tieteenfilosofioita. **Kielentutkimuksen tieteenfilosofiaksi** kutsutaan sitä periaatteiden ja käsitysten joukkoa, joka ohjaa kielentutkimuksen tekemistä.

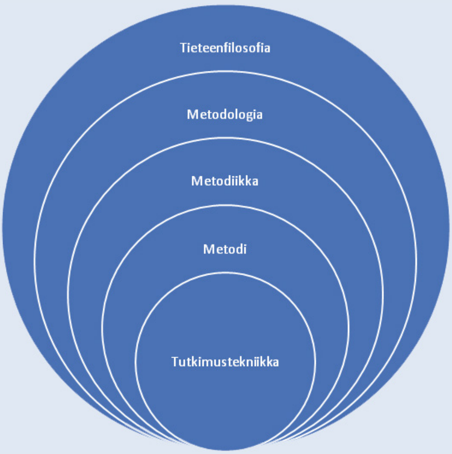
**Metodologia** eli **menetelmäoppi** tutkii metodien filosofisia ja loogisia perusteita. Metodologia on osa tieteenfilosofiaa, ja sillä tarkoitetaan

tieteellisen menetelmän määrittelemistä, kehittämistä ja soveltamista koskevien periaatteiden ja ehtojen tutkimusta. Metodologiassa tutkitaan siis tieteellistä menetelmää eli keinoja, joilla tieteellistä tietoa hankitaan. Metodologia tarjoaa siten välineitä menetelmällisten valintojen perusteluun.

**Metodiikka** on jollain tieteenalalla käytössä olevien tieteellisten menetelmien, tutkimustekniikoiden ja välineiden joukko. Metodiikka voidaan jakaa ainakin seuraavaan neljään osaan: aineiston keräämisen tekniikat, aineiston kuvaamisen tekniikat, aineiston analyysin tekniikat ja päätelmien tekemisen tekniikat. Metodologian ja metodiikan keskinäinen erossa pitäminen on sikäli hankalaa, että molemmat edellyttävät toisiaan.

**Metodi** eli **tutkimusmenetelmä** on tieteellinen tutkimusprosessi, joka johtaa tieteelliseen tietoon. *Metodeilla* viitataan niihin erityisiin keinoihin, välineisiin ja tutkimustekniikkoihin, joilla tutkimuskohteesta pyritään hankkimaan tieteellistä tietoa.

**Tutkimustekniikka** on tieteellisen tutkimusprosessin konkreettinen osa, jollaisia ovat esimerkiksi otantatekniikka, haastattelu- tai kyselytekniikka, numeeristen aineistojen muodostaminen ja erilaiset tilastolliset menetelmät.



**Teoria** on lauseiden joukko, joka muodostaa yhtenäisen kuvauksen jostakin ilmiöstä. Teorioita on eritasoisia. Kielentutkimuksessa yksittäisen teorian voi muodostaa vaikkapa joukko tietyn kielen kielioppia koskevia väitteitä, jotka muodostavat kielen teoreettisen osa- tai kokonaiskuvauksen. *Taustateorialla* viitataan puolestaan tutkimuksessa käytettyyn laajempaan teoriaan tai tietopohjaan, joka antaa kuvauksen kohteestaan ja kertoo kohteesta suhteessa muihin todellisuuden ilmiöihin.

## 1.2. Aiheen valinta ja tutkimuksen teoreettinen tausta

Tutkimusaiheen valinta ja alustava tutkimuskysymysten määrittely ovat tutkimuksenteon jännittävimpiä ja usein myös haastavimpia vaiheita. Tutkimusideat voivat kummuta esimerkiksi tutkijan omista kokemuksista ja aikaisemmista tutkimuksista, tai ne voivat tulla annettuina esimerkiksi tutkimusprojektissa, jonka osaksi tutkija liittyy. Hyvin tavallista on, että tutkija saa idean tutkimukselleen lukemalla muiden tutkijoiden aikaisemmin tekemiä tutkimuksia. Hän voi esimerkiksi havaita tutkimuksessa aikaisemmin kartoittamattoman aukon, jonka hän nyt haluaa selvittää. Tutkija voi myös pyrkiä testaamaan muiden tutkijoiden esittämien teorioiden toimivuutta erilaisissa aineistoissa.

Henkilökohtaisten tekijöiden lisäksi aiheen valintaan vaikuttavat myös yhteisölliset tekijät. Esimerkiksi kielentutkimuksessa eri tutkimusperinteillä on usein omat tapansa käsitteistää kieltä tutkittavana ilmiönä. Tällainen alustavakin tutkittavan ilmiön määrittely vaikuttaa siihen, millaiset tutkimusaiheet näyttäytyvät kyseisessä tutkimusperinteessä mielekkäinä ja mitkä eivät. Kielentutkimuksessa kohteena ei ole näin ollen kieli tai jokin yksittäinen osa kieltä sinänsä vaan kieli jostain näkökulmasta tarkasteltuna. (Ks. Longino 1990, 98–102.) Erilaisia näkökulmia kieleen ja kielellisten ilmiöiden tarkasteluun esitellään tämän kokoelman luvuissa 1–24.

Tutkimusidean hiominen varsinaiseksi tutkimusaiheeksi on tyypillisesti monivaiheinen prosessi. Hyvän tutkimusaiheen ominaispiirteenä

voidaan pitää ainakin sitä, että aihe on sellainen, että sitä ylipäänsä on mahdollista tutkia. Tämä tarkoittaa esimerkiksi sitä, että tutkimusaiheen kartoittamiseen vaadittavaa aineistoa on saatavilla ja sitä on mahdollista kerätä ja analysoida käytettävissä olevan ajan ja resurssien rajoissa. Etenkin aloittelevien tutkijoiden tutkimusaiheet ovat usein aluksi varsin laajoja ja osin vielä määrittymättömiä, ja niitä hiotaan täsmällisemmiksi tutkimuksen kuluessa, kun tutkijan ymmärrys aiheesta kasvaa. Tutkimusaiheen määrittelyssä myös muilta tutkijoilta ja mahdollisilta tutkimustyön ohjaajilta saatu palaute on usein arvokasta. Onkin hyvä huomata, että käytännön tutkimusprosessissa tutkimuksen eri vaiheet ovat aktiivisia samanaikaisesti eikä yhtä vaihetta suljeta pois, kun siirrytään seuraavaan. Esimerkiksi tutkimusaihe laajenee, rajautuu ja tarkentuu koko prosessin ajan. Aiheenvalinta on tavallaan valmis vasta, kun tutkimus on valmis.

Toinen hyvän tutkimusaiheen ominaispiirre on se, että aiheen tutkimista voidaan pitää jollain tavalla merkityksellisenä. Etenkin tutkimusrahoitusta hakiessaan tutkijan tulisi pystyä perustelemaan mahdollisimman selkeästi, miksi aihe on tutkimisen arvoinen. Perustutkimuksen osalta olennaista on osoittaa, kuinka tutkimus lisää tietoa ja ymmärrystä aihepiiristä. Tutkimuksella voi olla myös soveltavampaa yhteiskunnallista merkitystä. (Ks. Creswell 2014, 27; myös Johdannon luku 1.6. tk.)

Yksi tapa tarkastella tutkimuksen merkitystä on suhteuttaa se aikaisempiin tutkimuksiin. Tutkimuksen kirjallisuuskatsauksella on tässä keskeinen rooli, joten siihen kannattaa panostaa. Kirjallisuuskatsauksessa tutkija sitoo tutkimuksensa aiheesta käytyyn aikaisempaan tieteelliseen keskusteluun kertaamalla, millaisia tutkimuksia aiheesta on aikaisemmin tehty. Samalla hän voi osoittaa mahdollisia aukkoja ja puutteita aikaisemmassa tutkimuksessa sekä perustella, millaisia uusia metodologisia, teoreettisia ja käytännöllisiä näkökulmia hänen tutkimuksensa alalle tarjoaa. Hyvä kirjallisuuskatsaus ei siis tyydy luettelemaan aikaisemmin tehtyjä tutkimuksia vaan pohtii kriittisesti niiden vahvuuksia ja heikkouksia. Oman tutkimuksen suhteuttaminen aikaisempaan tutkimukseen on olennainen osa tieteellistä käytäntöä, ja se mahdollistaa tutkimuksen merkityksen arvioimisen. Yksittäisten tiedonmurusten merkitystä on vaikea arvioida, ellei niitä voida suhteuttaa mihinkään

aikaisempaan tutkimustietoon. Muiden tutkijoiden tekemän työn kunnioittaminen ja asianmukainen huomioiminen omassa tutkimuksessa ovat myös keskeinen osa hyvää tieteellistä käytäntöä (TENK 2012).

Kiinteä osa kirjallisuuskatsausta on tutkimuksen teoreettisen taustan määrittely. Tutkimuksen teoreettinen tausta kertoo, millaisessa käsitejärjestelmässä kohdetta on tulkittu. *Taustateorialla* viitataan tutkimuksessa käytettyyn teoriaan, joka antaa kuvauksen kohteestaan ja kertoo kohteesta suhteessa muihin todellisuuden ilmiöihin. Taustateoria sisältää ontologisia sitoumuksia, ja viime kädessä juuri näiden kohteen luonnetta koskevien oletusten yhteensopivuus metodologisten oletusten kanssa kertoo, onko menetelmällä saatu tieteellistä tietoa. Johdannon luvussa 2 käsitellään laajemmin sitä, mitä tarkalleen ottaen tarkoitetaan ontologisilla sitoumuksilla ja miten ne vaikuttavat metodologiaan. Yleistäen voidaan sanoa, että käsitykset siitä, mitä ja millaisia ilmiöitä maailmassa on, vaikuttavat käsityksiin keinoista saada tietoa noista ilmiöistä. Taustateoria ohjaa tutkimusta monilla tasoilla: mitä tutkimuskohteesta havainnoidaan, mitä tutkimuksessa pidetään tärkeänä ja mihin siinä kiinnitetään huomiota. Se tarjoaa käsitteitä, joilla kohdetta luokitellaan, ja ohjaa siten perustavalla tavalla tulosten tulkintaa. Samalla teoria rikastaa aiheesta käytyä pohdintaa sitomalla tutkimustulokset laajempaan kontekstiin, minkä ansiosta niiden arvoa on mahdollista tarkastella. Taustateoria tarjoaa siis tutkijalle välineitä käsitellä tutkittavaa ilmiötä systemaattisesti ja antaa selitysmalleja tutkimustulosten tulkintaan (selittämisestä ks. Johdannon luku 3 tk.; Eskola & Suoranta 1999, 80; ks. myös Creswell 2014, 51–76).

Teorioita käytetään laadullisessa ja määrällisessä tutkimuksessa (ks. Johdannon luku 1.3. tk.) kuitenkin usein hieman eri tavalla. Laadullisessa tutkimuksessa teoria yleensä tarjoaa tutkimukselle näkökulman, josta tutkimuskohdetta lähestytään, kun taas määrällisissä tutkimuksissa on yleisempää testata teorioita. Uusi teoria tai käsite voi olla myös laadullisen tutkimuksen tulos. (Ks. esim. Creswell 2014, 51–76; Luodonpää-Manni 2016.) Teorian erilaisista rooleista johtuu, että termillä *teoria* voidaan viitata myös hieman erilaisiin kohteisiin. Yhtäältä *teorialla* voidaan tarkoittaa ilmiön systemaattista esitystä (vrt. Kerlinger 1979, 64), joka on suhteellisen kattava, yhtenäinen ja vakiintunut tietorakennelma (esim.

prototyyppiteoria, puheaktiteoria, jokin yksittäinen kielioppiteoria jne.). Toisaalta *teoriaksi* tai *teoreettiseksi viitekehykseksi* voidaan kutsua tutkimuksen taustalla vaikuttavaa tietopohjaa, joka voi koostua eri lähteistä kerätyistä teoreettisista ja metodologisista periaatteista. Tällainen teoreettinen eklektismi on ollut tavallista suomalaisessa kielentutkimuksessa, jossa on vanhastaan usein vältetty sitoutumasta johonkin yksittäiseen teoriaan ja sen sijaan tutkimuksen teoriatausta on koostettu useista eri lähteistä. (Luodonpää-Manni 2016.) Eklektisen otteen omaksumisessa keskeistä on kuitenkin varmistaa, että taustateorioiden premissit, kategoriaoletukset ja loogiset implikaatiot ovat keskenään yhteensopivia. *Teoria* voi tarkoittaa kielentutkimuksessa myös yhden viitekehyksen mukaisesti laadittua kokonaiskuvausta kielestä (esim. systeemis-funktionaalinen kielioppi tai konseptuaalinen semantiikka). Tällaiset kuvaukset muodostuvat kokoelmasta teoreettisia väitteitä (esim. *-ssA* on suomen inessiivi), jotka ovat teoriapartikulaareja (ks. Johdannon luku 1.1. tk.).

Vaikka onkin mahdollista olla sitoutumatta mihinkään yksittäiseen teoreettiseen rakennelmaan, on hyvä huomata, että tutkijan aikaisemmat tiedot ja teoreettinen tausta ohjaavat havainnointia tarjoamalla käsitteet tutkittavan ilmiön tarkastelulle (Niiniluoto 1980, 224–225; ks. myös Johdannon luku 1.1. tk.). Esimerkiksi kieliopin ilmiöistä olisi vaikea puhua tukeutumatta aikaisemmassa tutkimuksessa käytettyihin käsitteisiin. Vaikka eri teorioissa käytettävät käsitteet olisivatkin lähellä toisiaan, niillä on joka tapauksessa historiansa ja käyttötapansa, jotka eivät välttämättä sovellu esimerkiksi uusien empiiristen löydösten kuvaamiseen suoraan. Kielenkuvaus johtaa näin välttämättä teoretisoimiseen. Lisäksi tutkimusmenetelmät, jopa tekniset työkalut, perustuvat teoreettisille olettamuksille, jotka ohjaavat sitä, millaisia tuloksia menetelmillä on mahdollista saada (esim. Latour & Woolgar 1986 [1979]). Kokenut tutkija tiedostaa omat taustaoletuksensa ja tuo ne tutkimuksessaan esille.

Meinefeldin (2004, 153–154) mukaan tutkijan aikaisemman tiedon vaikutusta havainnointiin voidaan pyrkiä kontrolloimaan ainakin kahdella eri tavalla. Aikaisemmat tiedot voidaan yhtäältä tehdä eksplisiittisiksi hypoteesien (ks. Johdannon luku 1.3. tk.) muodossa, jolloin aikaisemman tiedon vaikutuksia tutkimuksen kulkuun voidaan arvioida avoimesti. (Ks. myös Creswell 2014, 51–68.) Toisaalta aikaisemman

tiedon vaikutusta tutkimuksen kulkuun voidaan pyrkiä kontrolloimaan päinvastaisella strategialla, välttämällä tutkimusta ohjaavien, aikaisempaan tietoon perustuvien hypoteesien esittämistä. Tämä strategia on käytössä esimerkiksi ns. aineistovetoisissa tutkimuksissa (ks. Johdannon luku 3.3.2. tk.), joissa myös tutkimuskysymys usein johdetaan aineistosta. Tutkimustulokset eivät kuitenkaan nouse aineistosta itsestään, vaan kaikki tutkimus – niin teorialähtöinen kuin aineistolähtöinen – vaatii tutkijan tekemää tulkintaa (ks. Meinefeld 2004, 155). Tämä tulkinta ei tapahdu koskaan tyhjiössä, vaan siihen vaikuttaa esimerkiksi tutkijan saama koulutus.

Viime kädessä taustaoletusten rooli tiedon rakentumisessa ei kuitenkaan ole pelkästään yksittäisen tutkijan tai menetelmän vastuulla, vaan tiedeyhteisö määrittää lopulta sen, millaisia merkityksiä tieteelliseksi tiedoksi hyväksyttyihin väitteisiin jää ja mitkä niistä poistetaan: tieteen intersubjektiivisuus takaa sen, että lauseiden totuusarvo säilyy arvonäkökulmista huolimatta (Longino 1990, 216).

## 1.3. Tutkimusstrategia

Tutkimusaiheen valinnan ohella tutkimusstrategian määrittely on keskeinen osa tutkimusta pohjustavaa suunnitteluvaihetta. Tutkimusstrategia on yksittäistä metodia tai tutkimustekniikkaa laajempi menetelmällisten valintojen kokonaisuus, joka ohjaa tutkimuksen toteutusta. Tällaiset tutkimusstrategiset valinnat koskevat muun muassa sitä, käytetäänkö tutkimuksessa laadullisia, määrällisiä vai kenties monimenetelmäisiä lähestymistapoja, onko kyseessä vähän tutkittua aluetta kartoittava eksploratorinen tutkimus vai aikaisempia teorioita testaava konfirmatorinen tutkimus, onko kyseessä tapaustutkimus, vertaileva tutkimus jne. Tutkimusstrategiaan liittyvät menetelmälliset ratkaisut ovat tiukasti kytköksissä myös aineiston valintaan ja tutkimuskysymyksiin, sillä erilaiset menetelmät soveltuvat erilaisille aineistoille ja vastaavat erilaisiin tutkimuskysymyksiin (ks. Johdannon luku 3 tk.).

Laadullisissa eli kvalitatiivisissa tutkimusstrategioissa tutkimuskohdetta, sen ominaisuuksia ja merkityksiä pyritään ymmärtämään

kokonaisvaltaisesti. Se on omiaan aikaisemmin vain vähän kartoitettujen tutkimusalueiden tarkastelussa ja erityisesti silloin, kun se koskee ihmisen kokemusta ja sosiaalista toimintaa. Tutkimusstrategian kokonaisvaltaisen luonteen takia laadulliseen tutkimukseen soveltuvat usein pienemmät aineistot kuin määrälliseen tutkimukseen. Esimerkiksi tapaustutkimuksessa voidaan syventyä yksilön tai pienen ryhmän kielelliseen viestintään ja tuottaa näin näyttöä aikaisemmin muodostetun kieltä koskevan väitteen tueksi tai sitä vastaan. Laadullisessa tutkimuksessa ratkaisevaa ei ole aineiston (pieni) koko vaan niiden laadullisten tutkimuskysymysten luonne, joihin tutkimuksen kuluessa pyritään vastaamaan. Laadullisen tutkimuksen ominaispiirteitä ovat usein analyysin syvyys, rikkaus ja perusteellisuus, mikä lisää teoreettista ymmärrystä ilmiön luonteesta. (Laadullisesta tutkimuksesta ks. esim. Syrjäläinen, Eronen & Värri 2007.) Kielentutkimuksessa laadulliset menetelmät soveltuvat esimerkiksi kielellisten ilmiöiden kuvaukseen ja sen selvittämiseen, mikä kielessä ylipäänsä on mahdollista. Kielentutkimuksessa on perinteisesti käytetty laadullisia menetelmiä, jotka ovat perustuneet tutkittavan ilmiön luokitteluun ja sitä koskevien havaintojen tulkintaan aineistojen tai kirjallisuuden valossa. Teknologinen kehitys on kuitenkin helpottanut laajojen korpusaineistojen keräämistä ja käsittelyä (ks. Korpusaineistot tk.). Samalla kielentutkimuksen kohteita on alettu käsitellä yhä enemmän myös määrällisestä näkökulmasta.

Määrällisissä eli kvantitatiivisissa tutkimusstrategioissa tutkimuskohdetta ja sille ominaisia piirteitä pyritään selittämään mittaamisen avulla. Määrällisen tutkimuksen perustana on vahva teoria, joka mahdollistaa tarkasteltavien yksiköiden operationalisoinnin eli saattamisen mahdollisimman tarkasti mitattavaan muotoon (ks. Johdannon luku 1.5. tk.). Määrällinen tutkimus soveltuu laajojen aineistojen analyysiin ja erilaisten muuttujien välisten suhteiden tarkasteluun. Lisäksi voidaan tarkastella vaikkapa yksittäisen muuttujan ilmenemistä eri ryhmien välillä ja tuottaa näin tietoa, jota ei olisi voitu saada ilman ryhmien välistä vertailua. (Määrällisestä tutkimuksesta ks. esim. Töttö 2000.) Kielentutkimuksessa määrällisellä tutkimuksella selvitetään usein kielellisen ilmiön yleisyyttä ja säännönmukaisuuksia. Myös laadullisessa tutkimuksessa voidaan ilmaista määrällisiä tietoja. Määrien ilmaisu ei sinällään

vielä kerro, onko kyseessä määrällinen vai laadullinen tutkimus, vaan keskeistä on, onko tutkimuskysymys laadullinen vai yleisyyttä koskeva: johdetaanko yksittäistapauksista yleistyksiä (induktio) vai päätelläänkö mittaustuloksista ilmiöitä koskevia piirteitä (deduktio) (induktiivisesta ja deduktiivisesta päättelystä ks. tarkemmin Johdannon luku 1.5. tk.; Niiniluoto 1983).

Käytännöllisestä näkökulmasta määrälliset ja laadulliset menetelmät voidaan nähdä toisiaan täydentävinä lähestymistapoina (ks. Creswell 2014, 10–11). Usein näkökulmat vuorottelevat jopa samassa tutkimuksessa. Erilaiset tutkimusnäkökulmat rikastuttavat tietoa tutkimuskohteesta. Tutkimusnäkökulmia voidaan jakaa esimerkiksi sen mukaan, missä määrin ne objektivoivat kohdettaan tai missä määrin kohteen subjektiivinen ääni on tutkimuksessa esillä (ks. kuvio 2). Tutkijan on erilaisia metodologisia jakoja pohtiessaan arvioitava ainakin kahta asiaa: 1) mitä päämäärää jako palvelee, ja 2) mitä alaa ajatellen jako on alun perin tehty. Yleinen käsitys ihmistieteissä – joihin luetaan humanistis-yhteiskuntatieteelliset tutkimusalat laajasti ymmärrettynä – on, että tutkimuskohdettaan havainnoiva ja luokitteleva tutkija katsoo sitä ikään kuin ulkoapäin. Tällöin tutkija lähestyy kohdetta omine käsitteistöineen ja pyrkii (usein määrälliseen) esitykseen jonkin havaittavan ilmiön luonteesta. Laadullisessa ihmistieteellisessä tutkimuksessa voidaan vuorostaan pyrkiä löytämään kokonaan uusia merkityksellisiä ilmiöitä: voidaan esimerkiksi avoimien kysymysten avulla kysyä tietyltä ryhmältä heidän omia käsityksiään jostakin asiasta. Yllä olevassa jaossa on huomattavissa näkökulma ihmiseen. Kun ihmistä havainnoidaan ulkoapäin, hänellä ei ole valtaa vaikuttaa tutkimuskäsitteistöön. Kun ihmistä sitä vastoin haastatellaan ja kuunnellaan, hänen äänensä kuuluu. Kielentutkimuksessa tällaista jakoa voi problematisoida: Mitä tarkoittaa, että puhujan ääni kuuluu tai että kieltä objektivoidaan? Mitä päämääriä jako laadulliseen ja määrälliseen palvelee kielentutkimuksessa?

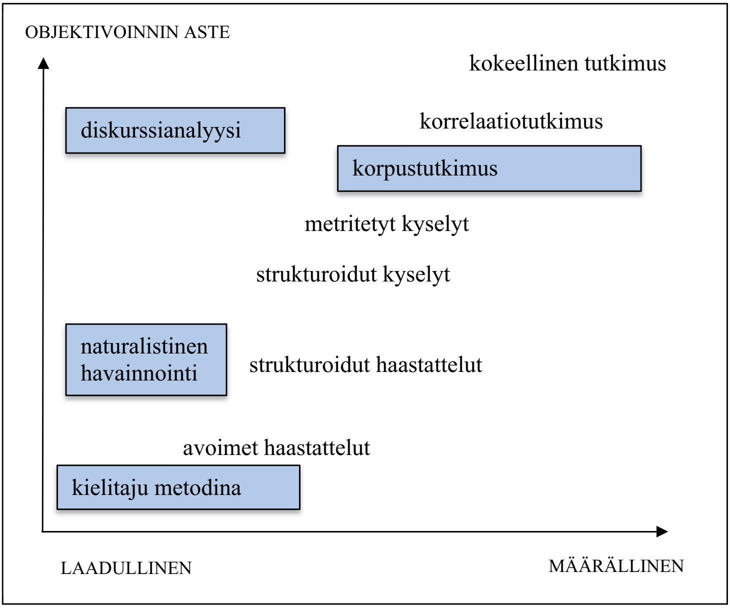
Kielentutkimuksessa laadullisessakin tutkimuksessa sovelletaan käsitteitä, luokituksia ja malleja. Vastaavasti määrällisessä tutkimuksessa, esimerkiksi suuria korpuksia tutkittaessa, on mahdollista, että löytyy uusia ilmiöitä. Keskustelu laadullisesta ja määrällisestä tutkimuksesta onkin hyvä ymmärtää karkeaksi yleistykseksi, joka edesauttaa

tietynlaisten erojen ymmärtämistä eri metodityyppien välillä. Jakoa ei tule ymmärtää kategorisena tai siten, että metodityyppien välille edes olisi tarvetta tehdä jonkinlainen tiukka raja. Metodit on hyvä mieltää erilaisina prototyyppisinä kategorioina, joiden kriteerit osa laadullisista tai määrällisistä metodeista täyttää paremmin ja osa heikommin. Jotkin kielentutkimuksessa käytetyt metodit ehkä jopa häilyvät määrällisen ja laadullisen tutkimuksen välillä. Vaikka yleensä on niin, että laadullinen tutkimus kuvaa, analysoi, määrittelee ja luokittelee ilmiötä ja määrällinen tutkimus vuorostaan selvittää yleisyyttä ja yleistettävyyteen liittyviä kysymyksiä, kielentutkimuksessa on kuitenkin myös uusia tutkimusaloja tai metodeja, jotka eivät sovi tähän jakoon kovin hyvin. Esimerkiksi typologiassa tutkitaan usein ilmiöiden yleisyyttä ja aineistona ovat usein kielioppikuvaukset, eli tutkimuksen päämäärä on yhteneväinen määrällisen ja aineisto laadullisen tutkimuksen kanssa.

Kielitieteen metodologiassa esiintyy myös toisenlaisia eroja suhteessa ihmistieteiden metodologian standardikäsitteistöön. Tutkimustyyppiä, jota humanistisilla aloilla usein kutsutaan fenomenologiseksi, on kielentutkimuksessa perinteisesti käsitteellistetty *intuitiona* tai *introspektiona* (ks. Kielitaju kielentutkijan työkaluna tk.). Kun tutkija nojaa omiin kokemuksiinsa, tietoonsa tai ajatuksiinsa ilmiöstä, esimerkiksi estetiikan tajuun tai kielitajuun, perspektiivi on subjektiivinen. Tällöin metodin tieteellisyys perustuu tutkijan käsitteelliseen kykyyn eritellä omia intuitioitaan sellaisella tavalla, että ne ovat tutkijayhteisölle ymmärrettäviä, jossakin taustateoriassa käsitteellistettyjä ja tuovat lisätietoa aiheesta suhteuttamalla niitä aikaisempaan tietoon. Pelkkä vapaa tajunnanvirta ilman intuitioiden suhteuttamista teoreettiseen keskusteluun ei millään alalla ole tiedettä.

Metodit voidaan siten jakaa sisäisen ja ulkoisen näkökulman, määrällisen ja laadullisen lähestymistavan sekä subjektiivisuuden ja objektiivisuuden asteen mukaan. Nuoren ja kehittyvän tieteenalan koulukuntakiistat johtuvat usein siitä, että jonkin näkökulman kannattajat uskovat, että yksi näkökulma on ylitse muiden tai että jokin uusi näkökulma ilmiöön joko on erityisen perusteltu tai ei ole perusteltu. Kypsä tieteenala pyrkii sen sijaan aktiivisesti kehittämään uusia metodeja, joilla voidaan rikastaa käsityksiä ilmiöistä. Se pohtii erilaisten metodologisten yläjakojen funktiota myös käytännön näkökulmasta. Kuvioon 2 on

sijoitettu perinteistä ihmistutkimusta koskevia erilaisia tutkimusasetelmia objektivoinnin asteen mukaan sekä laadullisen ja määrällisen akselilla. Kuvio 2 on sovellus Konsteniuksen (2014, 167) esittämästä kaaviosta, ja siihen on lisätty sinisellä pohjalla kielitieteelle tyypillisiä metodeja.




Kuvio 2. Laadullinen ja määrällinen tutkimus Konsteniusta (2014, 167) mukaillen.

Kielitieteessä yksi olennainen aineistonhankintamenetelmä on kielenkäytön tarkkailu eli naturalistinen havainnointi. Harjaantunut kielentutkija tulee huomanneeksi uusia ilmiöitä tai mielenkiintoisia piirteitä tutuissa ilmiöissä (esim. Karlsson 1998, 286–287). Esimerkiksi muutama vuosi sitten englannin kautta levisi suomeenkin totutusta poikkeava konjunktion käyttö: *vierailun aikana en ehdi sanoa yhtäkään lausetta loppuun, koska lapset ja väsymys ja kamala häslinki*. Konjunktiot aloittavat tavallisesti sivulauseen, mutta *koska x* -rakenteessa on usein vain substantiivilausekkeita. (Ilmiöstä ks. Sinnemaa 2015.) Myös erilaiset

lapsukset ja muut erikoisuudet kiinnittävät huomiota; niissä voi piillä vakiintumassa olevaa variaatiota ja jopa kielen muutoksen ituja. Esimerkiksi suomen potentiaali voi harvinaisuuttaan monelle äidinkielisellekin tuottaa vaikeuksia (*tulenee* pro *tullee*, *menenee* pro *mennee* jne.). Kielentutkimuksessa naturalistinen havainnointi voidaan joiltain osin tulkita laadulliseksi tutkimukseksi, sillä se harvoin on strukturoitua, objektivoivaa tai kovin teoriapitoista. Se on pikemminkin spontaania kielentutkijan omaan kielitajuun ja tietoon perustuvaa huomiointia luonnollisissa kielenkäytön tilanteissa. Havainnointi on metodina kuitenkin sikäli satunnainen, että ilmiöitä putkahtaa havaittaviksi ennakoimattomasti. Se on metodina kielitieteessä korvaamaton mutta vaatii tuekseen muita strukturoidumpia menetelmiä, jotta ilmiön yleisyydestä, laadusta ja ominaispiirteistä voisi saada tieteellisesti yleistettävää tietoa.

Tutkimuskohdetta voidaankin lähestyä myös monimenetelmäisesti. Monimenetelmäisiä tutkimusstrategioita on yleisesti ottaen kaksi: triangulaatio ja sarja-asetelma. **Triangulaatio** on monimenetelmäinen tutkimusstrategia, jossa saman tutkimuskohteen ja -kysymyksen tarkastelussa käytetään useita erilaisia tutkimusmenetelmiä. Triangulaation taustalla on ajatus, että kohdistamalla useita erilaisia tutkimusmenetelmiä samaan tutkimuskohteeseen ja kysymyksenasetteluun tutkimuskohteesta voidaan saada luotettavampia tuloksia kuin käyttämällä vain yhtä menetelmää, jos eri menetelmillä saavutetut tulokset ovat samansuuntaisia.

Toisessa monimenetelmäisessä tutkimusstrategiassa, **sarja-asetelmassa**, ensimmäisestä tutkimuskysymyksestä (k1) seuraa uusi tarkentava tutkimuskysymys (k2), johon seuraava tutkimus vastaa. Yleistäen voidaan sanoa, että siinä missä triangulaation avulla yhdestä ja samasta tutkimuskohteesta pyritään yleensä hankkimaan luotettavampaa kokonaisvaltaista tietoa, sarja-asetelman avulla tutkimuskohteesta pyritään hankkimaan monipuolisempaa tietoa kuin pelkästään yhtä tutkimusmenetelmää käyttäen olisi mahdollista saada. Sarja-asetelmassa yhdistetään usein laadullisia ja määrällisiä menetelmiä. Esimerkiksi aikaisemmassa laadullisessa tutkimuksessa havaitun ilmiön yleisyyttä voidaan pyrkiä selvittämään myöhemmässä määrällisessä tutkimuksessa, tai aikaisemmassa määrällisessä tutkimuksessa havaittuja kiinnostavia tai poikkeavia ilmiöitä voidaan tarkastella lähemmin ja selittää

myöhemmän laadullisen tutkimuksen avulla. Vaikka monimenetelmäisessä tutkimusstrategiassa yhdistäminen koskee usein määrällisten ja laadullisten menetelmien soveltamista samassa tutkimuksessa, on kuitenkin mahdollista yhdistellä myös esimerkiksi yksinomaan erilaisia laadullisia menetelmiä (ks. Creswell 2014; Konstenius 2014, 175–177).

Kielentutkimuksessa laadullisia ja määrällisiä menetelmiä yhdistelevä tutkimusstrategia (joskus jopa samassa tutkimuksessa) on osoittautunut hyvin hedelmälliseksi. Koska laadulliset ja määrälliset tutkimusasetelmat vastaavat erilaisiin tutkimuskysymyksiin, niitä yhdistelemällä tutkittavia ilmiöitä on mahdollista tarkastella eri näkökulmista. Näin tutkimuskohteesta saadaan monipuolista tietoa, mikä tekee oikeutta ihmiskielen monimuotoiselle olemukselle. Monimenetelmäisyyden hyötyjä kielentutkimuksessa ovat käsitelleet esimerkiksi Arppe ja Järvikivi (2007a; 2007b), Gries ja Divjak (2010) sekä Luodonpää-Manni, Penttilä ja Viimaranta (2017). Nykyisin tieteissä ja tiedepolitiikassa suositaan aloja ylittäviä monitieteisiä lähestymistapoja. Tämä näkyy suomalaisessa tiedepolitiikassa muun muassa Suomen Akatemian rahoittamina huippuyksikköinä, joilta tyypillisesti edellytetään monitieteisyyttä tai tieteidenvälisyyttä. Monitieteisessä tutkimuksessa samaa kohdetta tutkitaan useiden eri tieteiden omilla menetelmillä, kun taas tieteidenvälisessä tutkimuksessa yhdistetään eri tieteiden lähestymistavat ja luodaan uusi tapa lähestyä tutkimuskohdetta.

Aiheesta lisää:
Creswell, John W. 2014. *Research Design. Qualitative, Quantitative and Mixed Methods Approaches.* Los Angeles (CA): Sage.

## 1.4. Tutkimuksen aineistot ja eettiset kysymykset

Kuten edellisestä luvusta käy ilmi, aineiston laatu ja tarvittava koko ovat suoraan riippuvaisia valitusta tutkimusstrategiasta. Olipa valittu lähestymistapa laadullinen, määrällinen tai monimenetelmäinen, kielentutkijoiden saatavilla on merkittävä määrä erilaisia ja erikielisiä valmiita

aineistoja ja aineistolähteitä, joita voidaan soveltaa tutkimuksen tarpeisiin (ks. Korpusaineistot tk.). Valmiiden aineistojen hyödyntäminen on järkevää etenkin opinnäytetöitä suunniteltaessa, sillä se nopeuttaa tutkimusprosessia. Lisäksi valmiiden aineistojen kohdalla monet aineiston keräämiseen liittyvät kysymykset, kuten aineiston edustavuus ja luotettavuus sekä erilaiset laki- ja sopimusasiat, on jo ratkaistu. Kieliaineistoja ja niiden käsittelyyn sopivia työkaluja on saatavilla esimerkiksi kansallisessa FIN-CLARIN-konsortion tarjoamassa Kielipankissa[2] ja eurooppalaisessa CLARIN ERIC -tutkimusinfrastruktuurissa[3]. Tutkija on kuitenkin vapaa keräämään oman aineistonsa, jos oman tutkimuksen tarpeisiin soveltuvaa aineistoa ei ole valmiina tarjolla tai se on esimerkiksi joiltain osin riittämätöntä.

Kielentutkimuksessa käytettäviä aineistoja voidaan kerätä lukuisilla eri menetelmillä, joista monia esitellään tarkemmin tämän kokoelman artikkeleissa (mm. naturalistinen ja strukturoitunut havainnointi, kokeelliset tutkimusasetelmat, vuorovaikutustilanteiden tallentaminen, elisitaatio ja erilaisten tekstiaineistojen kokoaminen). Aineistoja voi kerätä tutkimustarkoituksiin suhteellisen vapaasti mistä tahansa julkaistuista teksteistä tai julkisista puheista. Sen sijaan yksityisten julkaisemattomien tekstien, esimerkiksi kirjoitelmien, kyselyvastausten, keskustelujen ja haastattelujen, keräämiseen tarvitaan tutkittavien suostumus. Suostumusta pyydettäessä on tärkeää kertoa tutkittaville helposti ymmärrettävällä tavalla, mistä tutkimuksessa on kysymys, sekä määritellä tarkasti, millaiseen tarkoitukseen aineistoa käytetään. Yleensä suostumusta ei kannata rajata vain yksittäiseen tutkimukseen, vaan se kannattaa pyytää esimerkiksi tieteelliseen tarkoitukseen tai tutkimus- ja opetuskäyttöön, jotta aineiston jälkikäteinen käyttö helpottuu ja on joustavaa. Suostumuspyynnön yhteydessä tutkittavia informoidaan myös heidän oikeuksistaan.

Aineiston keräämisessä on lisäksi noudatettava voimassa olevia tietosuojasäännöksiä[4]. Nyrkkisääntönä voidaan pitää sitä, että henkilötietoja kannattaa kerätä ja käsitellä mahdollisimman vähän. Henkilötiedoilla tarkoitetaan tässä esimerkiksi etu- ja sukunimiä, sähköposti- ja ip-osoitteita sekä muita henkilöön viittaavia tietoja, joiden perusteella yksityishenkilö on mahdollista tunnistaa (tapauskohtaisesti esim. käyttäjätunnukset, paikkakunnat, ammatit). Jos henkilötietojen käsittely ei

ole välttämätöntä tutkimuksen tulosten kannalta, aineisto on suositeltavaa kerätä nimettömänä. Tarvittaessa henkilötietoja voidaan kuitenkin kerätä tieteellistä tutkimusta varten tietosuojasäännösten puitteissa. EU:n tietosuoja-asetus (GDPR, General Data Protection Regulation) ja sitä täydentävä, vuoden 2019 alusta voimaan tullut tietosuojalaki noudattavat riskiperusteista lähestymistapaa, jossa tietosuojan edellyttämät menettelytavat mitoitetaan suhteessa arvioituun riskiin, joka yksityishenkilölle voi koitua henkilötietojen paljastumisesta. Henkilötietoja käsittelevän tutkijan tuleekin suunnitella henkilötietojen suojaustoimenpiteet koko käsittelyn elinkaareksi. Ne tulee pystyä myös osoittamaan tietosuojaviranomaiselle (henkilörekisterin tietosuojaperiaatteiden osoitusvelvollisuus). Lisäksi tietosuojarikkomuksesta, esimerkiksi henkilötietoja sisältävän muistitikun katoamisesta, on tehtävä ilmoitus tietosuojaviranomaiselle. Voimassa olevista tietosuojasäännöksistä saa lisätietoa Tietosuojavaltuutetun toimiston[5] lisäksi myös yliopistojen järjestämistä tietosuojakoulutuksista. Tekijänoikeudesta lisätietoa tarjoavat esimerkiksi Kopiosto[6] ja Sanasto[7].

Laki- ja sopimusasioiden ohella aineiston keräämisen periaatteita ohjaavat myös monet tiedepoliittiset linjaukset, jotka suosivat ns. **avoimen tieteen** periaatteita.[8] Avoimien julkaisujen lisäksi painotetaan yhä enemmän myös tutkimusaineistojen avaamista, minkä toivotaan tehostavan aineistojen jatkokäyttöä ja parantavan tutkimuksen läpinäkyvyyttä. Samalla julkaistuista tutkimusaineistoista on tullut varteenotettavia julkaisuja varsinaisten tutkimusjulkaisujen rinnalla. Kieliaineistoja voi avata esimerkiksi FIN-CLARINin kautta. Se huolehtii aineistojen pitkäaikaistallennuksesta ja tarjoaa tutkijoille apua aineiston luovutukseen liittyvissä kysymyksissä.

Tutkimusaineiston avaaminen edellyttää huolellista suunnittelua jo aineiston keräämisvaiheessa. Jotta aineisto olisi käyttökelpoinen, siihen tulee liittää erilaisia aineiston alkuperää ja sen keräämistä koskevia metatietoja (ks. Korpusaineistot tk.). Tällaisia tietoja ovat esimerkiksi tekstin lähde, kirjoittaja tai puhujat, tekstin laajuus ja tekstilaji sekä tekstin tuottamisen ajankohta. Lisäksi aineiston metatietoihin on syytä kirjata myös tiedot siitä, kuinka raaka-aineistoa on käsitelty, koodattu ja rakenteistettu, jotta se on saatu muutettua ja analysoitua varsinaiseksi

tutkimusaineistoksi. *Raaka-aineistolla* tai *raakadatalla* tarkoitetaan tässä siis sisältöä, jota ei ole koodattu tai luokiteltu määrällisesti tai laadullisesti tulkintaa varten. *Tutkimusaineistolla* tai *tutkimusdatalla* puolestaan tarkoitetaan analogisia ja digitaalisia resursseja, jotka on koodattu ja luokiteltu päätelmien tekemistä varten. *Datasta* puhutaan erityisesti digitaalisessa muodossa olevien aineistojen yhteydessä. Eron tekeminen raaka-aineiston ja tutkimusaineiston välille on tärkeää, jotta tutkijan rooli aineiston luokittelussa ja kuvaamisessa tulee paremmin näkyville. Tutkimustuloksia ei johdeta suoraan raaka-aineistosta, vaan usein tarvitaan erityisiä aineiston kategorisoimisen ja kuvaamisen tekniikoita (ks. Johdannon luku 1.5. tk.), jotta aineistosta voidaan tehdä päätelmiä. Kielentutkimuksessa tällaisia voivat olla vaikkapa fonemisointi, kieliopillisten kategorioiden tunnistaminen tai sana- ja lausesemanttinen analyysi, siis kielentutkimuksen tavanomainen analyysitoiminta. On hyvä huomata, että käytettävät tekniikat sisältävät teoreettisia valintoja, jotka vaikuttavat aineiston tulkintaan. Esimerkiksi suullisten aineistojen litteroinnissa valittu menetelmä vaikuttaa siihen, millä tarkkuudella aineistoa kuvataan, käytetäänkö foneettista transkriptiota vai ei, missä määrin ilmeitä ja eleitä otetaan huomioon ja kirjataanko tutkimusaineistoon myös ympäristön tapahtumia (ks. esim. Ochs 1979).

Koska aineistonhallintaan sisältyy monia ratkaistavia kysymyksiä, aineiston keräämiseen, käyttöön, säilytykseen ja jatkokäyttöön liittyvät periaatteet on hyvä avata erillisessä aineistonhallintasuunnitelmassa. Suunnitelma auttaa tutkijaa aineistoon liittyvässä tiedonhallinnassa, ja sitä edellytetään yhä useammin myös tutkimusrahoitusta haettaessa. Esimerkiksi Suomen Akatemia edellyttää hakijoilta aineistonhallintasuunnitelmaa, jossa eritellään, minkä tyyppisiä tutkimusaineistoja hankkeessa kerätään tai käytetään, miten tutkimusaineisto dokumentoidaan, miten aineisto tallennetaan, miten taataan sen tietoturva ja yksilönsuoja, miten aineiston omistus- ja käyttöoikeuksia hallinnoidaan sekä miten aineiston jatkokäyttö mahdollistetaan.[9] Aineistonhallintasuunnitelma ei kuitenkaan ole ainoastaan tutkimuksen alkuvaiheen suunnitteluun liittyvä apuväline, vaan se tulee pitää ajan tasalla tutkimuksen edistyessä. Aineistonhallintasuunnitelmien tekemiseen on tarjolla tutkijan työtä helpottavia työkaluja, kuten DMPTuuli.[10] Kuten kaikissa tutkimuksen

vaiheissa tutkijan tulee noudattaa myös aineiston keräämisessä tutkimuksen eettisiä periaatteita.[11] Johdatuksen tutkimusetiikkaan tarjoaa esimerkiksi Kuula (2006).

**Aiheesta lisää:**

Avoin tiede: https://www.avointiede.fi/.

Kuula, Arja. 2006. *Tutkimusetiikka. Aineistojen hankinta, käyttö ja säilytys*. Tampere: Vastapaino.

Tietoarkiston aineistonhallinnan käsikirja: http://www.fsd.uta.fi/aineistonhallinta/fi/.

Tietosuojavaltuutetun toimisto: https://www.tietosuoja.fi/.

## 1.5. Analyysi ja tulosten esittely

Analyysin tavoitteena on tavallisesti esittää vastauksia tutkijan esittämiin tutkimuskysymyksiin. Tutkimuskysymykset sisältävät usein sellaisia teoreettisia käsitteitä, jotka tulee operationalisoida eli saattaa sellaiseen muotoon, että niitä on mahdollista mitata tai havainnoida mahdollisimman tarkasti. Esimerkiksi lausuman pituus voidaan operationalisoida erilaisten teoreettisesti määriteltyjen yksiköiden kautta, kuten laskemalla foneemien, morfeemien, tavujen tai sanojen määriä (Stefanowitsch 2010, 359–360). Samalla tavalla esimerkiksi kielellinen merkitys voidaan operationalisoida hyvin eri tavoin, muun muassa psykolingvistisissä koeasetelmissa mittaamalla kognitiivista prosessointia vaikkapa silmänliiketutkimuksella tai reaktioaikatesteillä (ks. Psyko- ja neurolingvistiikka tk.), kollokaatiotutkimuksissa tutkimalla sanan suhdetta ympäröivään kielelliseen kontekstiin (ks. Määrällinen korpuslingvistiikka tk.) tai tarkastelemalla sanan viittaussuhdetta ulkoisen maailman tarkoitteeseen (esim. millaisia ominaisuuksia vaatekappaleella tulee olla, jotta sitä voidaan nimittää *leggingseiksi*) (Geeraerts, Grondelaers & Bakema 1994).

Tutkimuskysymysten selvittämisen apuna käytetään usein ilmiötä koskevia oletuksia eli hypoteeseja, joita on erilaisia. **Tutkimus-** tai **selityshypoteesi** on tieteellistä argumentaatiota noudattava selitysmalli, jolla ei ole vielä näyttöä tukenaan. Tutkija voi itse koetella hypoteesia, tai hän voi tarjota sitä muulle tiedeyhteisölle arvioitavaksi ja koeteltavaksi. Hypoteesi voi saada vahvistavaa näyttöä tuekseen, tai hypoteesia koetteleva tutkimus voi sen kumota. Tästä syystä tutkimushypoteesi on syytä muotoilla siten, että sen testaaminen on ylipäänsä mahdollista. Tutkimushypoteesit perustuvat tavallisesti ilmiöstä aikaisemmin tehtyihin havaintoihin tai tutkimuksiin, mutta ne voivat olla myös luovia arvauksia. Testattava tutkimushypoteesi olisi esimerkiksi kielenkäyttöä koskeva väite, että tietynlainen, negatiivissävyinen naisten miehille kohdistama puhe nimetään helpommin *nalkuttamiseksi* kuin vastaavantyyppinen puhe miehiltä naisille (ks. Vanhatalo 2005). Hypoteesia voisi testata korpusaineistolla tai kyselyllä. Testaamista varten tutkimushypoteesi muokattaisiin operationaaliseen muotoon. Kummassakin tutkimustyypissä tutkijan olisi pohdittava, millä lailla puheen suunta sukupuolelta toiselle ja sukupuoli itse operationalisoitaisiin tutkimusasetelmassa. Korpustutkimuksessa luonnollisen kielen luonnehdinta ”puhe nimetään helpommin” käännettäisiin määrällisesti mitattavaksi. Kyselytutkimuksessa samaa väitettä selvitettäisiin henkilöiden intuitioiden voimakkuudella ja mitattaisiin kyselyn vastausvaihtoehtojen skaalojen avulla.

Operationalisoinnin ja hypoteesin välillä on tärkeä ero. Yleisenä periaatteena on, että tulisi välttää perustamasta metodia tai tutkimusmallia väitteille, joilla on pelkkä hypoteesin asema eikä ollenkaan todistusaineistoa tukenaan. Nykytietämyksen ja -teknologian valossa vaikeasti testattava on esimerkiksi kognitiivisessa kieliopissa esitetty väite, että kielenkäyttäjät prosessoivat aivoissaan finiittiverbit ja ei-finiittiset verbit eri tavalla. Karkeistaen edelliset hahmotettaisiin elokuvamaisesti, jälkimmäiset valokuvamaisesti. (Ks. Langacker 1987, luku 7.) Tällä hetkellä meillä ei ole keinoja tutkia, simuloivatko ihmiset eri verbimuodoilla kuvattuja tilanteita valokuvien tai elokuvien kaltaisin kognitiivisin prosessein mielessään. Kyseessä on kuitenkin empiirinen väite, mikä tarkoittaa, että sitä periaatteessa voisi testata. Jos meillä olisi (tai joskus tulevaisuudessa on) käytössä ihmisten mielikuvia havainnoiva teknologia,

voisimme ratkaista empiirisen evidenssin valossa, pitääkö Langackerin (mp.) esittämä teoreettinen väite paikkansa.

Universaalisten eksistenssilauseiden testattavuudesta on käyty tieteenfilosofiassa paljon keskustelua. Nämä lauseet ovat muotoa ”kaikki joutsenet ovat valkoisia” tai ”puhe-elimillä tuotetun ihmiskielen keskeiset piirteet, kuten lineaarisuus ja kaksoisjäsennys, ovat minkä tahansa kielen universaaleja piirteitä”. Emme voi tutkia kaikkia menneitä ja tulevia joutsenia tai tietää, onko ollut olemassa tai tuleeko olemaan muita kuin valkoisia joutsenia (on esimerkiksi käynyt ilmi, että Australiassa on musta laji, *Cygnus atratus*). Emme myöskään voi tietää, onko havaittavassa maailmankaikkeudessa ollut olemassa tai tuleeko olemaan jokin oliolaji, jonka kieleksi luokiteltava vuorovaikutusjärjestelmä poikkeaa tyystin ihmiskielestä. Tällaisilla universaaleilla lauseilla ei tieteessä voi loogisesti olla evidenssiä oikeutuksenaan, ja ne tulee siksi tutkimuksessa muotoilla rajatuiksi eksistenssilauseiksi, esimerkiksi ”Euroopan mantereen joutsenilla on geneettinen perimä, joka tuottaa valkoisia fenotyyppejä” tai ”puhe-elimillä tuotetun kielen keskeiset piirteet, kuten lineaarisuus ja kaksoisjäsennys, ovat ihmiskielille tyypillisiä”. Rajattuja eksistenssilauseita voidaan yleensä testata. Yksittäistä tutkimusta varten väitteen rajausta supistetaan edelleen tutkimuksen empiiristä aineistoa vastaavaksi. Jos aineistona on esimerkiksi Kauhavalla puhuttava eteläpohjalainen nykymurre, tämä toimii tutkimuksen eksistenssiväitteiden rajauksena.

Tutkimushypoteesi ei ole sama asia kuin tutkijan yksityinen havainto tai tutkijan ennakko-oletus tutkimuksen kohteen luonteesta. Jälkimmäinen on kaikilla eri alojen tutkijoilla välttämättä oltava: tutkija tarvitsee ainakin jonkinlaisen ennakko-oletuksen siitä, millaista ilmiötä hän tutkii, jotta hän voi ylipäätään rakentaa tutkimusasetelman kokeillakseen ideoitaan. Tällaista ennakko-oletusta kutsutaan **työhypoteesiksi**. Tutkimushypoteesista poiketen työhypoteesia ei ensisijaisesti pyritä testaamaan tutkimuksen kuluessa, vaan se hyväksytään alustavasti tutkimusta ohjaavaksi periaatteeksi. Työhypoteesi koostuu teoreettisista ennakko-olettamuksista, jotka mahdollistavat sen, että tutkija voi muodostaa alustavan käsityksen tutkittavasta ilmiöstä. Jo aineiston kerääminen vaatii usein ainakin jonkinlaisen ennakko-oletuksen tutkittavasta ilmiöstä.

Tutkijan ei kuitenkaan tarvitse kiinnittyä tutkimuksen alussa muotoilemaansa työhypoteesiin, vaan sitä voidaan korjata tutkimuksen kuluessa, mikäli ennakko-olettamukset osoittautuvat epätarkoiksi tai vääriksi.

Tutkimus- tai selityshypoteesin ja työhypoteesin eroa voi havainnollistaa seuraavalla esimerkillä. Kielentutkijalla on kielenkäytön tarkkailun kautta saatu havainto Y (tietyissä suomen morfeemeissa ei odotuksenvastaisesti toteudukaan rajageminaatio, esim. *hernekeitto* lausutaan [hernekeitto], ei [hernek:eitto]) jostakin kielen yleisestä ilmiöstä Z (suomen tiettyjä morfologisia muotoja koskeva rajageminaatio- tai loppukahdennusilmiö). Kielentutkija muotoilee työhypoteesin T (mitkä tekijät vaikuttavat siihen, että rajageminaatio on vähenemässä tai häviämässä tietyistä muodoista). Tämän perusteella hän ideoi sellaisen aineisto-otannan A (rajageminaatiomuotoja sisältävää puhetta Suomen kattavalla murreotannalla), jonka hän uskoo vakuuttavan muun tiedeyhteisön kielenilmiön olemassaolosta. Hänellä voi myös olla selityshypoteesi S (miksi tietyistä muodoista rajageminaatio on vähentynyt tai




Kuvio 3. Esimerkki työhypoteesin ja selityshypoteesin tehtävistä kielentutkimuksessa.

häviämässä), joka kertoo, miksi ilmiö Z toimii juuri niin kuin toimii. Onnistuneimmillaan tutkija osaa rakentaa työhypoteesinsa perusteella sellaisen tutkimusasetelman, jossa kaikki havainnot Y {1...n} ilmiöstä Z aineistossa A on selitettävissä S:n avulla. Työhypoteesi on siis sellainen odotuksista koostuva väite (valistunut arvaus), joka ohjaa tutkimuksen kulkua ja jonka tutkimustulokset voivat osittain vahvistaa tai kieltää.

Tutkimustulosten valmistuttua tutkijan tulee vielä arvioida kriittisesti niiden luotettavuutta. Luotettavuuden arvioinnissa esimerkiksi operationalisoinnilla sekä erilaisilla metodien uskottavuutta (robustisuus, tarkkuus, poikkeavien havaintojen merkitys) arvioivilla menetelmillä on keskeinen merkitys. Voihan olla, että tulokset olisivat paremmin selitettävissä jollain toisella hypoteesilla tai että hypoteesin testaamisessa käytetty tieteellinen menetelmä ei soveltunutkaan tarkoitukseen parhaiten. Virhepäätelmien poissulkemiseksi tilastollisessa testaamisessa käytetään yleensä ns. nollahypoteesiasetelmaa. Tällöin tutkimuksessa asetetaan tutkimushypoteesin lisäksi sille vastakkainen **nollahypoteesi** ilmaisemaan sellaista odotettavissa olevaa tulosta, jossa kahden ilmiön välillä ei ole yhteyttä. Lääketieteessä nollahypoteeseja käytetään osoittamaan, että jokin hoito on tehoton. Kielentutkimuksessa nollahypoteeseja voidaan käyttää tasapainottamaan päätelmän oikeutusta. Jos aineistossa ei ole riittävästi näyttöä nollahypoteesin hylkäämiselle, riski tehdä väärä päätelmä on yleensä suuri. Nollahypoteesia käytetään siis kilpailevana hypoteesina. Jos sitä ei voida perustellusti hylätä, ei voida olla varmoja, että tutkimushypoteesin muotoilemien ilmiöiden välillä todella olisi syy-yhteys tai muu kytkös. On tärkeää huomata, että nollahypoteesi ilmaisee jonkin ilmiön poissaoloa, eikä sitä siten voi koskaan loogisesti todentaa eli verifioida. Nollahypoteesi voidaan ainoastaan kumota eli falsifioida tai jättää kumoamatta. (ks. Johdannon luku 2.5. tk.)

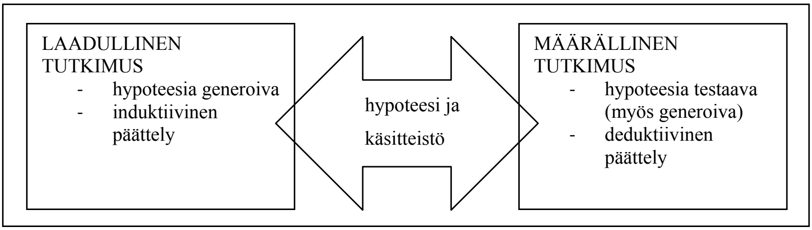
Tutkija voi nojata tutkimuksessaan erilaisiin päättelytapoihin: induktiiviseen, deduktiiviseen tai abduktiiviseen päättelyyn. **Induktiivinen päättely** lähtee liikkeelle käytännöstä ja etenee aineistossa esiintyvien yksittäistapausten tarkastelusta kohti havaintojen perusteella tehtyjä yleisempiä päätelmiä. Induktiivinen päättely soveltuu esimerkiksi aikaisemmin vain vähän tutkittujen alueiden eksploratoriseen tutkimukseen. Tällöin tutkimushypoteeseja ei välttämättä muodosteta, mutta

erityisesti laadullisissa tutkimuksissa voidaan hyödyntää alustavasti hyväksyttyjä työhypoteeseja, jotka toimivat tutkimuksen teoreettisena kehyksenä. Laadulliset tutkimukset perustuvat tyypillisesti induktiiviseen päättelyyn, jossa tutkija johtaa yksittäistapauksista yleistyksiä. Havainto yksittäistapauksista perustuu aineistoon, joka on ensin luokiteltu jonkin taustateorian käsittein ja luokituksin. Rajatusta aineistosta ei yleensä voida induktiivisesti johtaa koko populaatiota eli perusjoukkoa koskevia päätelmiä. *Perusjoukolla* viitataan niihin tutkimusyksiköihin, joista tutkimuksella halutaan sanoa jotain. Esimerkiksi yksittäisiä suomen kielen puhujia havainnoimalla ei loogisesti voida tehdä kaikkia suomen kielen puhujia koskevia päätelmiä. Sen sijaan havaintojen perusteella on mahdollista muodostaa eli generoida hypoteesi, jota voidaan testata toisella päättelytavalla, esimerkiksi deduktiivisesti.

**Deduktiivinen päättely** etenee induktioon nähden päinvastaisessa järjestyksessä, toisin sanoen yleistyksestä yksityistapausta koskevaan päätelmään. Deduktiivinen päättely soveltuu esimerkiksi aikaisempia teorioita tai hypoteeseja testaavaan konfirmatoriseen tutkimukseen. Määrälliset tutkimukset perustuvat tyypillisesti tällaiseen deduktiiviseen päättelyyn. On kuitenkin huomattava, että myös määrällinen tutkimus voi olla hypoteesia generoivaa ja näin ollen eksploratiivista eli tutkimusaluetta kartoittavaa. Käytännössä voidaan lähteä liikkeelle teoriasta tai aikaisemman tutkimuksen tuloksista ja testata näistä johdettujen hypoteesien paikkansapitävyyttä yksittäistapauksissa. Jos esimerkiksi aikaisemmassa haastattelututkimuksessa huomataan, että monet haastateltavista assosioivat *kissa*-sanaan yllättävällä tavalla feminiinisen sukupuolen, tästä voidaan generoida hypoteesi ”*kissaan* liitetään suomessa feminiininen sukupuoli”. Hypoteesia voidaan sitten testata otoksen avulla (ks. Kysely tk.), vaikkapa poimimalla puhelinluettelosta joka 2 000:s nimi kyselyn pohjaksi. Määrällisessä tutkimusasetelmassa osallistujilta kysytään tämän jälkeen tyypillisesti suljettuja kysymyksiä, joiden määrää ja laatua mitataan. Tällaista päättelymuotoa kutsutaan **hypoteettis-deduktiiviseksi**, sillä rakenteellisesti tutkimusasetelma perustuu kahden lauseen väliseen suhteeseen, hypoteesiin ja tulokseen. Tutkimuksen tulos joko kieltää tutkimushypoteesin tai antaa esitetylle hypoteesille tukea.

Induktion ja deduktion lisäksi tutkijan tekemät päätelmät voivat perustua myös abduktiiviseen päättelyyn. **Abduktiivinen päättely** hyödyntää sekä taustateoriasta johdettuja hypoteeseja että aineistosta tehtyjä havaintoja ja etenee todennäköisimpään selitykseen. Abduktiossa itse päättelyprosessi voi edetä kumpaan tahansa suuntaan eli havainnoista teoriaan tai teoriasta havaintoihin. Ilmiötä koskevien päätelmien lisäksi tutkija voi siis kehittää tutkimustulostensa perusteella myös uusia hypoteeseja ja teorioita, joita voidaan sitten testata tulevissa tutkimuksissa.

Kuvio 4 kiteyttää edellä kuvatut laadulliseen ja määrälliseen tutkimukseen liitettävät perusominaispiirteet.




Kuvio 4. Laadullinen ja määrällinen tutkimus karkeasti esitettynä monimenetelmäisessä tutkimusasetelmassa (Konstenius 2014, 174).

Aineiston analysoinnissa käytettävät tarkemmat metodit ovat suoraan sidoksissa valittuun tutkimusstrategiaan. Siinä missä määrällisissä tutkimuksissa aineistoa analysoidaan usein erilaisia tilastollisia menetelmiä hyödyntäen, laadullisissa tutkimuksissa aineistoa voidaan jäsentää esimerkiksi taustateorioihin perustaen. Jokainen tutkimus sisältää tyypillisesti ainakin jonkinlaisen aineistonkeruumenetelmän, jolla tutkimusaineisto on hankittu, sekä tutkimusmenetelmän (tai -menetelmiä), jonka avulla aineistosta on pyritty hankkimaan uutta tietoa. Aineiston keräämisen tekniikoita ovat esimerkiksi havainnointi, koeasetelmat, kyselyt, haastattelut ja muu (kieli)aineistojen kerääminen. Aineiston kategorisoimisen ja kuvaamisen tekniikoita ovat muun muassa aineiston luokittelu, korpusaineistojen annotointi ja suullisten aineistojen litterointi ja transkribointi. Aineiston analyysin tekniikoita ovat esimerkiksi kieliopillisten kategorioiden tunnistaminen, sana- ja lausesemanttinen

analyysi, erilaiset tilastolliset ja matemaattiset menetelmät sekä taustateoriasta kumpuavien käsitteiden hyödyntäminen aineiston jäsennyksessä. Lisäksi voidaan erottaa erityisiä päätelmien tekemisen tekniikoita (ks. Johdannon luku 2.6. tk.). Tutkimuksen eri vaiheisiin liittyvien tekniikoiden erittely voi helpottaa oman tutkimuksen jäsentämistä ja lisätä näin myös tutkimuksen tarkkuutta. Erittelyn apuna voidaan käyttää myös johdannon luvussa 2 esitettyjä metodin ehtoja.

Sekä laadullisissa että määrällisissä tutkimuksissa aineistoa voidaan esitellä kokoavasti taulukoiden ja kuvioiden avulla tai esimerkinomaisesti lainaamalla näytteitä alkuperäisaineistosta. Aineiston sisällön julkaisemisessa on kuitenkin huolehdittava siitä, että yksityishenkilöitä ei voida tunnistaa näytteen perusteella. Tämä tarkoittaa sitä, että viimeistään tässä vaiheessa aineisto on anonymisoitava eli siitä on poistettava tai muutettava kaikki henkilöön viittaavat ja muut yksilöivät tiedot. Tämänkin jälkeen aineistoon voi joissain tapauksissa jäädä sellaista arkaluontoista materiaalia, jonka kohdalla tutkija voi joutua pohtimaan, missä määrin aineistoa on tarpeen julkistaa. Tutkijan tulee noudattaa tutkimustyössä hyvän tieteellisen käytännön periaatteita, kuten rehellisyyttä, yleistä huolellisuutta ja tarkkuutta (TENK 2012). Tämä koskee tutkimuksen koko elinkaarta: tutkimuksen suunnittelua, varsinaista tutkimustyötä, tutkimustulosten julkaisemista ja tutkimustulosten ja aineistojen jatkokäyttöä. Vastuullinen tiede -sivusto tarjoaa lisätietoa tutkimusetiikan ja tiedeviestinnän hyvistä käytänteistä tutkimuksen elinkaarimallin jokaisessa vaiheessa.[12]

**Aiheesta lisää:**

TENK 2012. Hyvä tieteellinen käytäntö ja sen loukkausepäilyjen käsitteleminen Suomessa. Tutkimuseettisen neuvottelukunnan ohje 2012. Helsinki: Tutkimuseettinen neuvottelukunta. http://www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf.

Vastuullinen tiede. Tutkimusetiikka ja tiedeviestintä Suomessa: https://www.vastuullinentiede.fi/.

## 1.6. Miten kielentutkimuksen menetelmillä voidaan vastata ihmiskunnan suuriin haasteisiin?

Kielellä on keskeinen rooli kaikessa yksilön ja yhteiskunnan psykososiaalisessa toiminnassa: sen avulla hahmotetaan maailmaa, vaihdetaan ajatuksia ja tehdään yhteiskunnallisia päätöksiä. Koska kielenkäyttö ja kielelliset sopimukset ohjaavat ihmisyhteisöjen toimintaa monin tavoin, ihmisyhteisöissä havaitut ongelmat ovat usein sidoksissa kieleen ja kommunikaatioon (Tiililä 2017a; 2017b; ks. myös Searle 1995; 2010). Kielenkäyttöön liittyviä ongelmia ei kuitenkaan tunnisteta kovin hyvin. Ihmiskunnan suurista haasteista puhuttaessa katse käännetään usein yhteiskuntatieteilijöihin tai luonnontieteilijöihin, jotka myös hakevat kielentutkijoita aktiivisemmin valtion strategista tutkimusrahoitusta, vaikka strategisen tutkimuksen painopisteissä on myös kielentutkijoille sopivia teemoja.[13] On tärkeä ymmärtää, että koska kieli on keskeinen osa kaikkea yhteiskunnallista toimintaa, kielentutkija tutkii yhteiskuntaa suoraan ja hänellä on siten paljon mahdollisuuksia vastata yhteiskunnan ja ihmiskunnan suuriin haasteisiin. Erilaisia listauksia näistä haasteista ovat tehneet muun muassa Suomen Akatemia, Unescon Millenium-hanke[14] ja YK[15]. Haasteet ovat luonteeltaan sellaisia, että kielentutkimus tai mikään muukaan tieteenala ei yksin pysty ratkaisemaan niitä kaikkia, vaan tarvitaan tieteidenvälistä yhteistyötä. Kielentutkijoilla on kuitenkin käytössään välineitä ja osaamista, jotka voivat toimia tärkeänä osana haasteiden ratkaisemisessa.[16]

Kielentutkijoiden ydinosaamista ovat kielenkäyttö, ihmisten välinen vuorovaikutus ja kulttuurienvälinen viestintä, joihin monet maailman syvimmistä ongelmista tavalla tai toisella liittyvät. Kielentutkijoita tarvitaan tunnistamaan näitä kieleen ja kommunikaatioon liittyviä haasteita. Tällä hetkellä akuutteja ovat esimerkiksi globalisaation, digitalisaation ja eriarvoistumisen kysymykset, joihin kaikkiin voidaan löytää ratkaisuja tavoitteellisen dialogin avulla. Kielentutkijat voivat edesauttaa keskinäistä ymmärrystä ja selvittää, miten esimerkiksi verkkoviestintä muuttaa ihmisten välistä kommunikaatiota ja yhteiskuntaa. Yksi hyvä sovellusesimerkki tästä on tekoälytutkija Timo Honkelan (2017) rauhankonehanke, jossa keskeisessä roolissa on kielellinen kommunikaatio ja

tekoälyn käyttäminen ihmisten keskinäisen ymmärryksen lisäämisessä. Radikalisoitumista pyritään ratkomaan viestinnällisin keinoin monissa hankkeissa (ks. mm. IC Thinking Cambridge; Suomessa Helsingin kaupungin hanke ”Minun silmin, sinun silmin”[17]). Kieliteknologian, kuten puheenkäsittelyn, puheentunnistuksen ja puhesynteesin, kehitys tarjoaakin paljon mahdollisuuksia merkityksellisen vuorovaikutuksen lisäämiselle.

Keskinäinen ymmärrys on keskeistä myös kulttuurien vuoropuhelua koskevissa haasteissa. Kun ratkaistaan erilaisia konflikteja ja globaalin muuttoliikkeen mukanaan tuomia ongelmatilanteita, erilaisten näkökulmien ymmärtäminen on keskeisellä sijalla. Kielentutkimuksen keinoin voidaan parantaa edellytyksiä kulttuurien vuoropuhelulle ja helpottaa esimerkiksi maahanmuuttajien ja muun väestön sopeutumista muuttuvaan tilanteeseen pyrkimällä lisäämään ymmärrystä ja suvaitsevaisuutta (ks. Heikkilä & Niiniluoto 2016; Luodonpää-Manni ja muut 2017). Keskinäinen ymmärrys ja vuorovaikutus ovat myös liberaalin demokratian kulmakiviä. Kielentutkijat voivat omalla panoksellaan tuoda esiin myös kielivähemmistöjä ja uhanalaisia kieliä koskevia kysymyksiä sekä vaikuttaa näiden tukemiseen. Tässä työssä on tärkeää muun muassa kielien dokumentointi ja kuvaaminen sekä harvinaisten kielten revitalisaatio (ks. Kielten dokumentointi ja kieliopin kuvaus; Antropologinen lingvistiikka tk.).

Yhteiskunnan näkökulmasta kielenkäyttöön liittyvien ongelmien parempi tunnistaminen olisi merkittävää esimerkiksi yhteiskunnan eriarvoistumiskehityksen katkaisemiseksi. Koska julkinen valta konkretisoituu pitkälti kielen avulla, teksteillä on keskeinen rooli julkishallinnon työssä. On viitteitä siitä, että etäännyttävä ja ulossulkeva virkakieli johtaa julkisten palveluiden alikäyttöön ja on osaltaan vaikuttamassa esimerkiksi nuorten syrjäytymiseen (Tiililä 2017a; 2017b). Tämän vuoksi monet puhuvat mieluummin syrjäyttämisestä kuin syrjäytymisestä. *Syrjäytyminen* terminä jättää vastuun yksilölle itselleen, kun taas *syrjäyttäminen* nostaa esiin yhteiskunnallisten rakenteiden (ja kielenkäytön) vaikutuksen tässä prosessissa. Syrjäytyminen ja syrjäyttäminen toimivat samalla esimerkkinä siitä, miten kieli on aina näkökulmaista. Kielellinen tuotos ei ole neutraali, vaan tällaisia merkitysneuvotteluja tulisi

käydä paljon useammin kuin nykyisessä julkisessa keskustelussa tavataan tehdä. Kielentutkimus tarjoaa tarkat käsitteet ja metodit tällaisten kielellisten merkitysneuvotteluiden purkamiseen.

Kielentutkijoita tarvitaan myös suomen kielen kehittymisen tarpeisiin. Tieteen kielen painottuessa yhä enenevässä määrin englantiin tarvitsemme suomen kielestä ja sen käytöstä tutkimusta, jota myös julkaistaan suomeksi. Jos suomenkielisen terminologian kehittäminen laiminlyödään, voi olla, että pian ei ole edes mahdollista käydä tieteellistä keskustelua suomeksi. Tällöin riskinä on suomen sulkeutuminen tieteellisen kielenkäytön ja tiedediskurssin ulkopuolelle. Eräs tällaista kehitystä torjuva hanke on Tieteen termipankki, jonka tavoitteena on koota ja ylläpitää kaikkia tieteenaloja koskevaa avointa termitietokantaa tutkijoiden ja kansalaisten käyttöön. Verkkopalvelu[18] avattiin vuonna 2012, ja sitä ylläpidetään talkooperiaatteella eri alojen tutkijoiden kanssa. Kielitiede on ollut yksi ensimmäisistä Tieteen termipankin pilottihankkeeseen osallistuneista tieteenaloista.

Kielentutkijoilla on annettavaa sellaisiinkin globaaleihin haasteisiin kuin ilmastonmuutokseen vastattaessa. Vaikka ilmastonmuutosta sinänsä ei voida suoraan pysäyttää kielentutkimuksen menetelmillä, on huomattava, että ratkaisuja ilmastonmuutosta koskeviin ongelmiin haetaan kielelliseen kommunikaatioon ja vuorovaikutukseen perustuvien poliittisten neuvotteluiden kautta. Esimerkiksi ilmastosopimusten aikaansaaminen vaatii laajaa yhteisymmärrystä ilmastonmuutoksen luonteesta, seurauksista ja keinoista, joilla sitä voidaan torjua. Yhteisymmärryksen puute puolestaan voi vakavasti vaarantaa tai jopa estää globaalien ongelmien ratkaisut. Koska kieli ei ainoastaan kuvaa vaan myös luo todellisuutta, ilmastonmuutoksen ja muiden globaalien haasteiden olemassaolosta ja sen kiistämisestä käytäviä merkitysneuvotteluja voidaan purkaa ja tuoda näkyväksi kielentutkimuksen keinoin. Globaaleihin haasteisiin vastattaessa olennainen edellytys yhteisymmärryksen syntymiselle on lisäksi kulttuurienvälinen kommunikaatio, joka on niin ikään kielentutkimuksen ominta alaa.

Koska tutkimusten mukaan humanistien, siis myös kielentutkijoiden, asiantuntemusta arvostetaan yhteiskunnassa (Heikkilä & Niiniluoto 2016), tutkijoiden tulisi entistä aktiivisemmin ottaa osaa

yhteiskunnalliseen keskusteluun ja saattaa tuloksiaan eri sidosryhmien saataville. Tuomas Heikkilän ja Ilkka Niiniluodon (2016) tutkimuksessa kansanedustajien tuntemus humanistisesta tutkimuksesta oli eri päättäjäryhmien heikoin, ja kuitenkin juuri he tekevät isoja päätöksiä tieteen voimavarojen suuntaamisesta sekä ihmisiin ja yhteiskuntaan vaikuttavista poliittisista ratkaisuista. Kirjoittajien mukaan humanistien pitäisikin aktiivisemmin tiedottaa omista tutkimuksistaan myös akateemisen maailman ulkopuolelle. Tutkimuksen sisällön tuntemus lisää Heikkilän ja Niiniluodon tulosten mukaan myös sen arvostusta. Samalla voidaan antaa työkaluja poliittiseen päätöksentekoon.

Kieli koetaan usein hyvin henkilökohtaiseksi ja tunteita herättäväksi asiaksi, ja jos kielentutkijat eivät ota rooliaan alan asiantuntijoina, keskustelu jää vähemmän asiantuntevien keskustelijoiden ja usein liian tunnevaltaisen argumentoinnin varaan. Relevanttien tutkimustulosten tuottaminen ja niiden saattaminen ymmärrettävään muotoon ei kuitenkaan ole yksinkertaista, vaan siihen täytyy erikseen panostaa. Yliopistot tarjoavatkin tutkijoille säännöllisesti lisäkoulutusta tiedeviestinnässä. Samalla kuitenkin kielentutkijoilla on velvollisuus suhtautua kriittisesti siihen, millaista tiedeviestinnän tulisi olla, sillä sekin on kielellistä toimintaa.

Vaikka kielentutkimuksen menetelmillä on suuria yhteiskunnallisia sovellusmahdollisuuksia, samalla on kuitenkin syytä muistaa, että soveltavan tutkimuksen lisäksi kielentutkimuksessa tarvitaan myös sen omista lähtökohdista käsin määriteltyä perustutkimusta. Heikkilän ja Niiniluodon (2016, 82) mukaan ”[t]iedettä ovat aina vieneet eteenpäin suuret peruskysymykset, jotka ovat syntyneet tieteen sisältä, aiemman tutkimuksen avaamista ongelmanasetteluista”. Perustutkimus tuottaa tieteenalalle tukevan tietoperustan, jonka varaan soveltava tutkimus voi sitten rakentaa. Toisaalta soveltava tutkimus voi myös tarjota uusia perustutkimuksellisia kysymyksiä. Perus- ja soveltava tutkimus muodostavatkin jatkumon, jonka kaikkia sävyjä tutkimusyhteisössä ja yhteiskunnassa tarvitaan.

## VIITTEET

1 Voidaan ajatella, että kielen perustava kaksoisjäsennys on tunnettava ennen kuin aukeaa lingvistisen kuvauksen mahdollisuus kielen ”kolmannelle” jäsennykselle eli semioottisen (esim. peirceläisen) merkitsemistapahtuman analysoinnille.
2 https://www.kielipankki.fi.
3 https://www.clarin.eu/.
4 EU:n tietosuoja-asetuksen soveltaminen 25. toukokuuta 2018 alkaen on tuonut uusia oikeuksia omiin henkilötietoihin. Ks. tarkemmin https://www.tietosuoja.fi.
5 https://www.tietosuoja.fi/.
6 https://www.kopiosto.fi/.
7 https://www.sanasto.fi/.
8 https://www.avointiede.fi/.
9 https://www.aka.fi/tutkimusrahoitus/hae-rahoitusta/nain-haet-rahoitusta/ohjehakemisto/aineistonhallinta/aineistonhallintasuunnitelma/.
10 https://www.dmptuuli.fi/.
11 https://www.tenk.fi/.
12 https://www.vastuullinentiede.fi.
13 https://www.aka.fi/strateginen-tutkimus/.
14 https://www.un.org/sustainabledevelopment/.
15 http://www.un.org/en/sections/issues-depth/global-issues-overview/.
16 Kielentutkija Martin Hilpert käsittelee Youtube-videossaan kielentutkimuksen roolia liittyen yhteiskunnan suuriin haasteisiin: https://www.youtube.com/watch?v=X4OaN39sNAI&feature=youtu.be.
17 https://misi.hel.fi/.
18 http://tieteentermipankki.fi/wiki/Termipankki:Etusivu.

## KIRJALLISUUS


Arppe, Antti & Järvikivi, Juhani. 2007a. Take empiricism seriously! In support of methodological diversity in linguistics. *Corpus Linguistics and Linguistic Theory* 3:1, 99–109.

Arppe, Antti & Järvikivi, Juhani. 2007b. Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3:2, 131–159.

Creswell, John W. 2014. *Research Design. Qualitative, Quantitative and Mixed Methods Approaches*. Los Angeles (CA): Sage.

Eskola, Jari & Suoranta, Juha. 1999. *Johdatus laadulliseen tutkimukseen*. Tampere: Vastapaino.

Geeraerts, Dirk, Grondelaers, Stefan & Bakema, Peter. 1994. *The Structure of Lexical Variation. Meaning, Naming, and Context*. Berlin: Mouton de Gruyter.

Gries, Stefan Th. & Divjak, Dagmar. 2010. Quantitative approaches in usage-based Cognitive Semantics: Myths, erroneous assumptions, and a proposal. Julkaisussa: Glynn,

Dylan & Fischer, Kerstin (toim.) *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*. Berlin: Mouton de Gruyter, 333–353.

Harris, Zellig S. 1957 [1951]. *Methods in Structural Linguistics*. Chicago (IL): The University of Chicago Press.

Heikkilä, Tuomas & Niiniluoto, Ilkka. 2016. *Humanistisen tutkimuksen arvo. Kuusi murrettavaa myyttiä ja neljä uutta avainta*. [verkkoaineisto]. [viitattu 3.2.2018]. Helsinki: Opuscula Instituti Romani Finlandiae V. Saatavissa: http://irfrome.org/humanistisen-tutkimuksen-arvo/.

Honkela, Timo. 2017. *Rauhankone. Tekoälytutkijan testamentti*. Helsinki: Gaudeamus.

Karlsson, Fred. 1975. Fennistiikan tieteenparadigmasta ja sen ohjausvaikutuksesta. *Virittäjä* 79:2, 179–193.

Karlsson, Fred. 1998. *Yleinen kielitiede*. Uudistettu laitos. Helsinki: Yliopistopaino.

Kerlinger, Fred N. 1979. *Behavioral Research: A Conceptual Approach*. New York (NY): Holt, Rinehart, and Winston.

Kiikeri, Mika & Ylikoski, Petri. 2004. *Tiede tutkimuskohteena. Filosofinen johdatus tieteentutkimukseen*. Helsinki: Gaudeamus.

Konstenius, Reetta A. 2014. *Empiria, eksperimentti ja etiikka. Kielitieteen metateoriaa*. Helsinki: Helsingin yliopisto.

Kuhn, Thomas. 1994 [1962]. *Tieteellisten vallankumousten rakenne*. [*The Structure of Scientific Revolutions*.] Kääntänyt Kimmo Pietiläinen. Juva: WSOY.

Kuula, Arja. 2006. *Tutkimusetiikka. Aineistojen hankinta, käyttö ja säilytys*. Tampere: Vastapaino.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Volume I. Theoretical Prerequisites*. Stanford (CA): Stanford University Press.

Latour, Bruno & Woolgar, Steve. 1986 [1979]. *Laboratory Life: The Construction of Scientific Facts*. Princeton (NJ): Princeton University Press.

Longino, Helen E. 1990. *Science as Social Knowledge. Values and Objectivity in Scientific Inquiry*. Princeton (NJ): Princeton University Press.

Luodonpää-Manni, Milla. 2016. *'Théorie', 'hypothèse', 'modèle' et 'méthode' dans le domaine de la linguistique : Perspectives sémasiologique et onomasiologique*. Väitöskirja. Turku: Turun yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-29-6490-1.

Luodonpää-Manni, Milla, Penttilä, Esa & Viimaranta, Johanna. 2017. Introduction. Julkaisussa: Luodonpää-Manni, Milla, Penttilä, Esa & Viimaranta, Johanna (toim.) *Empirical Approaches to Cognitive Linguistics: Analyzing Real-Life Data*. Newcastle Upon Tyne: Cambridge Scholars Publishing, 1–21.

Luodonpää-Manni, Milla, Borg, Kaj, Nikanne, Urpo, Raitaniemi, Mia & Taivalkoski-Shilov, Kristiina. 2017. Kielitietoisuus, kielten ja kielentutkimuksen yhteiskunnallinen merkitys. [verkkoaineisto]. [viitattu 20.8.2018]. *Kieli, koulutus ja yhteiskunta* 9/2017. Saatavissa: https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-syyskuu-2017-1/kielitietoisuuden-kautta-kohti-ihmiskunnan-suuria-kysymyksia.

Martinet, André. 1964. *Elements of General Linguistics*. Kääntänyt Elisabeth Palmer. London: Faber and Faber Limited.

Meinefeld, Werner. 2004. Hypotheses and prior knowledge in qualitative research. Julkaisussa: Flick, Uwe, Kardorff, Ernst von & Steinke, Ines (toim.) *A Companion to*

*Qualitative Research*. [*Qualitative Forshchung – Ein Handbuch*.] Kääntänyt Bryan Jenner. London: Sage publications, 153–158.
Määttä, Urho. 1994. *Funktionaalinen selittäminen morfologiassa. Metateoriaa ja huomioita suomen ja sen sukukielten tutkimusperinteestä*. Opera Fennistica & Linguistica 7. Tampere: Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos.
Nida, Eugene. 1949. *Morphology. The Descriptive Analysis of Words*. (2. p.). Ann Arbor (MI): The University of Michigan Press.
Niiniluoto, Ilkka. 1980. *Johdatus tieteenfilosofiaan. Käsitteen- ja teorianmuodostus*. Helsinki: Otava.
Niiniluoto, Ilkka. 1983. *Tieteellinen päättely ja selittäminen*. Helsinki: Otava.
Ochs, Elinor. 1979. Transcription as theory. Julkaisussa: Ochs, Elinor & Schiefflen, Bambi B. (toim.) *Developmental Pragmatics*. New York (NY): Academic Press, 43–72.
Ojutkangas, Krista, Larjavaara, Meri, Miestamo, Matti & Ylikoski, Jussi. 2009. *Johdatus kielitieteeseen*. Helsinki: WSOY oppimateriaalit.
Percival, Keith. W. 1976. The applicability of Kuhn's paradigms to the history of linguistics. *Language* 52:2, 285–294.
Pike, Kenneth L. 1966 [1947]. *Phonemics. A Technique for Reducing Languages to Writing*. Ann Arbor (MI): The University of Michigan Press.
Sampson, Geoffrey. 1980. *Schools of Linguistics*. Stanford (CA): Stanford University Press.
Searle, John. 1995. *The Construction of Social Reality*. New York (NY): The Free Press.
Searle, John. 2010. *Making the Social World: The Structure of Human Civilization*. Oxford: Oxford University Press.
Sinnemaa, Tiina. 2015. *Joululaulut soi ja lapset pukee jo haalareitaan #koska #lumi*. Tutkimus *koska* x -rakenteen tulkinnasta. Pro gradu -tutkimus. Tampere: Tampereen yliopisto.
Stefanowitsch, Anatol. 2010. Empirical cognitive semantics: Some thoughts. Julkaisussa: Glynn, Dylan & Fischer, Kerstin (toim.) *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*. Berlin: Mouton de Gruyter, 355–380.
Syrjäläinen, Eija, Eronen, Ari & Värri, Veli-Matti. 2007. Johdanto. Julkaisussa: Syrjäläinen, Eija, Eronen, Ari & Värri, Veli-Matti (toim.) *Avauksia laadullisen tutkimuksen analyysiin*. Tampere: Tampere University Press, 7–12.
TENK 2012. Hyvä tieteellinen käytäntö ja sen loukkausepäilyjen käsitteleminen Suomessa. Tutkimuseettisen neuvottelukunnan ohje 2012. Helsinki: Tutkimuseettinen neuvottelukunta. [verkkoaineisto]. [viitattu 13.8.2018]. Saatavissa: http://www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf.
Tiililä, Ulla. 2017a. Johdatus sosiaali- ja hoitoalojen tekstimaailmaan ja niiden tutkimukseen. Julkaisussa: Tiililä, Ulla & Karvinen, Kati (toim.) *Elämän ja kuoleman tekstit: kirjoittaminen sosiaali- ja hoitotyössä*. Helsinki: Kotimaisten kielten keskus, 6–18.
Tiililä, Ulla. 2017b. Sosiaali- ja terveyspolitiikkaa kielen ja kielikäsitysten varassa. *AFinLAn Vuosikirja* 75, 54–71. [verkkoaineisto]. [viitattu 26.5.2018]. Saatavissa: https://journal.fi/afinlavk/article/view/60873.
Töttö, Pertti. 2000. *Pirullisen positivismin paluu. Laadullisen ja määrällisen tarkastelua*. Tampere: Vastapaino.
Vanhatalo, Ulla. 2005. *Kyselytestit synonymian selvittämisessä: sanastotietoutta kielenpuhujilta sähköiseen sanakirjaan*. Väitöskirja. Helsinki: Helsingin yliopisto.

# 2. Tieto, tieteellinen tieto ja tieteellinen menetelmä

*Reetta Konstenius*

*Markus Hamunen*
https://orcid.org/0000-0003-2786-6392

Tämän luvun tarkoituksena on selkeyttää tieteellisen metodin ehtoja siten, että lukijan tiedot ja ymmärrys metodien tieteellisyyden arvioinnista ja uusien luovien metodien kehittämisestä kasvavat. Luvussa kartoitetaan tieteellisten tutkimusmetodien jaettuja ominaisuuksia, ehtoja ja yleisiä periaatteita. Termiä *tieteellinen menetelmä* tai *metodi* käytetään niistä keinoista, joiden systemaattisen soveltamisen tuloksena saadaan tieteellistä tietoa. Luvussa toisin sanoen selvennetään sitä, mikä erottaa tieteellisen tutkimusmetodin muista vaihtoehtoisista tiedonhankintatavoista, tieteen näennäistieteestä ja tieteellisen tiedon pelkästä uskomuksesta. Luvun tavoite on jäsentää kielentutkimuksen metodologian yleisiä piirteitä siten, että se edesauttaa erilaisten tiedonhankintatekniikoiden kriittistä arviointia, innovatiivista kehittämistä ja yhdistämistä.

Tämä teos kokonaisuudessaan kuuluu sekä kielentutkimuksen metodologian että metodiikan alaan, sillä kirjan tavoitteena on antaa laaja-alainen kuva kielentutkimuksessa käytössä olevista tutkimustekniikoista. Johdannon tämä luku on metodologiaa (ks. myös Johdannon luku 3 tk.), sillä se vastaa kysymykseen, millä yleisillä kriteereillä jotakin tiedonhankintakeinoa pidetään tieteellisenä metodina. Metodologia tarkastelee myös tieteen metodiikkaa, sillä osa metodologisesta tutkimuksesta kohdistuu konkreettisiin, yksittäisiin käytössä oleviin metodeihin. Suuri osa tieteen metodeista kehittyy eri alojen arkitutkimuksen

lomassa, ei niinkään ylhäältä alas, teoriasta käytäntöön. Metodologiassa tutkitaan sitä, täyttävätkö nämä arkisen tutkimuksen myötä kehitetyt tutkimustekniikat tieteellisen metodin kriteerit. Koska metodologia kattaa sekä tutkimusmetodien teorian (normatiivisen aspektin) että niiden tutkimuksen (faktuaalisen aspektin), metodologiasta voidaan käyttää myös nimitystä *tutkimustekniikoiden metateoria*. Tämä metanäkökulma säilyy koko tämän teoksen kaikissa luvuissa: kielentutkimuksen metodeja kuvataan sekä metodiikan (sovelletut tutkimusmetodit) että metodologian (teoreettinen ylätaso) kannalta.

Tässä luvussa käsitellään tieteellisen metodin ehtoja. Johdannon luvussa 3 vuorostaan käsitellään laajemmin varsinaisten kielentutkimuksessa käytettyjen metodien teoriasidonnaisuutta ja kielentutkimuksen kenttää. Tämän luvun tavoite on auttaa lukijaa ymmärtämään, millaisista prosesseista voi käyttää termiä *metodi*. Luvun 3 tavoite on vuorostaan auttaa ymmärtämään metodin ja taustateorian suhdetta. Taustateoria kertoo, millaisia ominaisuuksia tutkimuskohteella on ja mitkä ovat ne tietoteoreettiset oletukset, jotka ohjaavat tutkimusmetodin tai -tekniikan kohdistamista kohteeseen.

## 2.1. Tieteellisen menetelmän ehdoista

Tieteellistä metodia pidetään yleensä sinä kriteerinä, jonka perusteella tiede voidaan erottaa muusta inhimillisestä toiminnasta: tieteellinen metodi erottaa tieteen ei-tieteestä. Tieteellinen metodi ja tiede ovat siten käsitteellisesti toisistaan riippuvaisia. Ne esiintyvät lähes aina toistensa määritelmissä ja implikoivat toinen toistaan, joskin eri tieteenfilosofisissa koulukunnissa eri tavoin. Jotkin koulukunnat painottavat kuitenkin enemmän tieteellistä prosessia ja toiset taas prosessin tuloksena saatavaa tietoa.

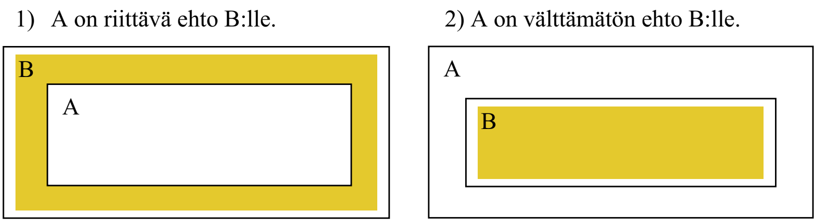
Mitä sitten tarkoittaa se, että tieteellinen metodi ja tiede implikoivat toisiaan? Implikaatio on looginen suhde etujäsenen ja takajäsenen välillä, missä edellinen edellyttää jälkimmäisen. Implikaatiota tarkastellaan riittävien ja välttämättömien ehtojen kautta. Väite A on B:n riittävä ehto, jos A:n totuus takaa sen, että myös B on tosi. Väite A on B:n välttämätön ehto, jos A ei voi olla tosi ilman, että B on tosi.

Riittävä ja välttämätön ehto implikaation avulla lausuttuna.

1) A on riittävä ehto B:lle, A → B

2) A on välttämätön ehto B:lle, B → A tai ei-A → ei-B

Riittävä ja välttämätön ehto voidaan esittää myös Venn-diagrammina kuten kuviossa 5.




Kuvio 5. Riittävä (1) ja välttämätön (2) ehto diagrammimuodossa.

Riittävillä ja välttämättömillä ehdoilla on metodologiassa kahtalainen sovellusalue. Toisaalta riittävien ja välttämättömien ehtojen avulla tutkitaan käsitteiden määritelmien suhteita – esimerkiksi juuri käsitteiden *tiede* ja *tieteellinen metodi* välistä suhdetta. Riittääkö esimerkiksi tieteellisen metodin käyttö määrittämään prosessin tieteeksi? Onko metodi tieteellisen tiedon välttämätön ehto? Toisaalta riittävien ja välttämättömien ehtojen avulla tarkastellaan kausaatiota eli predikaattilogiikassa jos–niin-rakennetta. Usein tieteessä pyritään selvittämään, mikä aiheuttaa jonkin tietyn ilmiön tai minkä toisen ilmiön kanssa tutkittava korreloi.

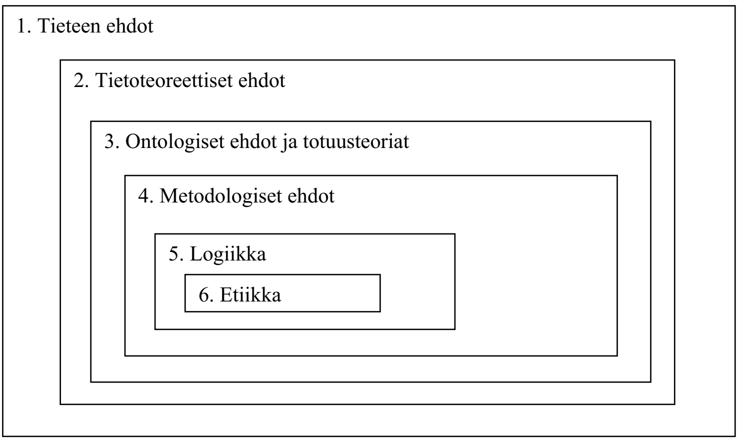
Välttämätön tarkoittaa sitä, että jokin B ei voisi tapahtua tai olla olemassa ilman, että on A. Painovoima on sunnuntaikävelyn välttämätön ehto, kuten kieli on puheen. Kuitenkin myös moni muu seikka vaikuttaa loppujen lopuksi siihen, että joku lähtee sunnuntaikävelylle ja puhuu. Riittävien ehtojen joukko on yleensä välttämättömiä ehtoja suurempi. Metodologiaan perehtyvän opiskelijan onkin mielekästä tutustua logiikan ja tieteellisen argumentaation perusteisiin, sillä metodin teoreettinen peruste on viime kädessä sen loogisessa rakenteessa, joka muodostaa tieteellisen argumentin rakenteen (esim. Allwood, Anderson & Dahl 1980 [1971]; Niiniluoto 1983).

**Aiheesta lisää:**

Kakkuri-Knuuttila, Marja-Liisa (toim.). 2000. *Argumentti ja kritiikki*. Helsinki: Gaudeamus.

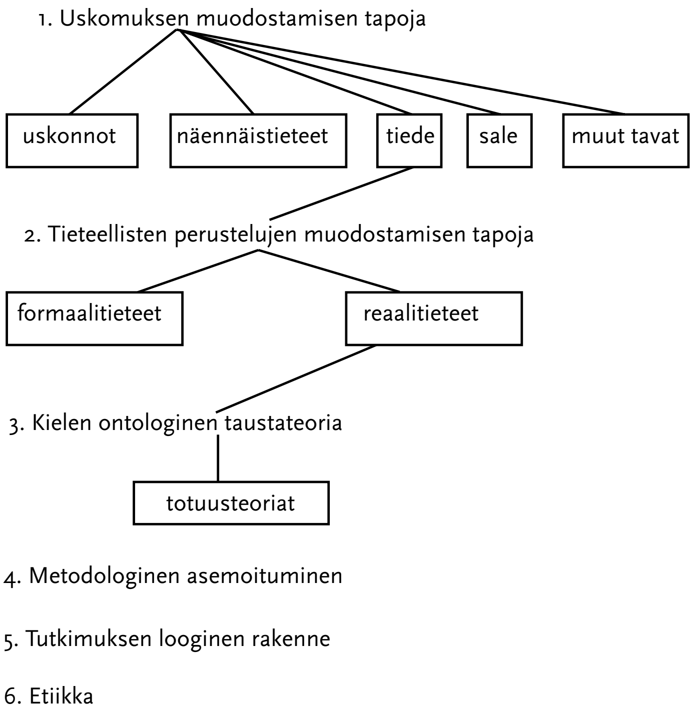
Miettinen, Seppo K. 1993. *Logiikan peruskurssi*. Helsinki: Gaudeamus.

Tieteellisen metodin määritelmä on siis loogisesti sidoksissa tieteen määritelmään. Seuraavaksi käsitellään tieteellisen metodin eräitä välttämättömiä ehtoja, jotka voidaan esittää kuvion 6 tapaan. Kuviossa ehtoja esitetään olevan kuusi, mutta jaottelu voisi olla myös toisenlainen. Ehtojen keskinäinen suhde perustuu ensinnä juuri välttämättömyyteen: ehto 1 on välttämätön ehdolle 2 jne. Näin ehdot kytkeytyvät toisiinsa. Toiseksi ehdot rakentuvat jossain määrin yleisemmästä yksityisempään -periaatteen varaan. Esimerkiksi tieteen ehdot (ehto 1) kuvaavat yleisiä tieteellisen tiedon edellytyksiä, kun taas esimerkiksi logiikassa (ehto 5) ja etiikassa (ehto 6) korostetaan yleisten seikkojen lisäksi myös tutkimuskohtaisia valintoja ja piirteitä (esim. mikroetiikka).




Kuvio 6. Tieteellisen menetelmän ehdot.

Ehtojen suhteet voidaan kuvata laatikkomallin sijaan myös puukuviona kuten kuviossa 7. Puumalli mahdollistaa tieteen ehtojen esittämisen havainnollisesti suhteessa muihin uskomuksen muodostamisen tapoihin.




Kuvio 7. Tutkimuksen teoreettiset ja toiminnalliset ehdot puukuviona.

Seuraavissa luvuissa tieteellisen menetelmän ehdot esitellään kohta kohdalta alkaen tieteen ehdoista ja edeten järjestyksessä aina kuudenteen ehtoon eli etiikkaan asti.

## 2.2. Tieteen ehdot: uskomuksen muodostamisen tapa

Tiede voidaan määritellä tavaksi muodostaa hyvin perusteltuja uskomuksia. Termillä *tiede* tarkoitetaankin yleensä joko tieteellisen tiedon tarkoituksenmukaista ja järjestelmällistä tavoittelua tai tieteen metodeilla saavutettujen uskomusten systemaattista kokonaisuutta. Samalla termillä viitataan siis usein niin tutkimuksen tuloksiin kuin tuloksiin johtaneeseen tieteelliseen toimintaan, tutkimusprosessiin.

Yleisen käsityksen mukaan tieteellisen tiedon subjekti on tiedeyhteisö, ei yksittäinen tutkija (Niiniluoto 1980a). Ei siis riitä, että kielentutkija on tehnyt mielestään kelpo tutkimuksen ja pitää tuloksiaan tosina; tiedeyhteisön on hyväksyttävä tutkimuksen tulos. Tieteessä tutkimuksen tulokset julkaistaan ja tutkimusprosessi asetetaan alttiiksi kriittiselle vertaisarvioinnille. Vasta kun tiedeyhteisö vakuuttuu väitteiden todenmukaisuudesta ja siitä, että tutkimuksessa käytetyt keinot ovat luotettavia, tutkimus hyväksytään tieteeksi.

Tätä seikkaa voi havainnollistaa seuraavan naturalistiseen havainnointiin eli kielenkäytön tarkkailuun liittyvän esimerkin avulla. Kielentutkija huomaa, että kahvilakeskustelussa *eiku*-partikkelia ei käytetä ainoastaan korjauksen keinona, toisin kuin kielitieteessä on aiemmin uskottu. Kutsutaan tätä uutta havaittua ilmiötä Y:ksi. Kielentutkija tallentaa keskusteluaineistoa, jonka hän myöhemmin litteroi. Analysoituaan aineiston kielentutkija uskoo keksineensä *eiku*-partikkelin käytön selityksen Z. Yksityinen havainto ilmiöstä ja selityshypoteesin kehittäminen eivät kuitenkaan ole vielä tieteellistä tietoa. Tieteellistä tietoa arvioidaan erityisin kriteerein, joista tieteellisen metodin käyttäminen on yksi ja tutkimuksen julkaiseminen toinen. Merkitystä on myös sillä, mitä julkaistaan ja missä. Vertaa esimerkiksi seuraavien julkaisujen näyttöä.

1) Kielentutkija julkaisee havainnon Y, ja muu tiedeyhteisö hyväksyy, että tällainen havainto Y on tehty.
2) Kielentutkija julkaisee havainnon Y ja selityksen Z, ja muu tiedeyhteisö hyväksyy esitetyn selityksen validiksi eli edelleen koeteltavaksi tiedeyhteisössä.

3) Kielentutkija julkaisee havainnon Y, minkä lisäksi hän kerää aineiston A. Aineistossa A selityksen Z voidaan osoittaa pitävän paikkansa kaikissa A:ssa esiintyvissä yksittäistapauksissa Y, ja tämän tiedeyhteisö hyväksyy.
4) Kielentutkija julkaisee havainnon Y, minkä lisäksi hän kerää keskenään erilaiset aineistot A, B ja C. Aineistoissa A, B ja C selitys Z kattaa kaikki Y:n yksittäisesiintymät, ja tämän tiedeyhteisö hyväksyy.

Esimerkkien 1–4 avulla huomataan, että tieteellinen tieto liittyy sitä tukevaan näyttöön. Kun ilmiötä edustava evidenssipohja eli aineisto on monipuolista ja esitetyt argumentit sekä tutkimuksen looginen rakenne pätee koko aineistoon, on tiedeyhteisön hyväksymiskynnys todennäköisemmin ylitettävissä. Esimerkeissä voi myös vaihdella tiedeyhteisön merkitystä ja pohtia, millä lailla tieto on (koko) tiedeyhteisön hyväksymä, jos sen julkaisee esimerkiksi laitoksen omissa julkaisuissa, väitöskirjana, posteriesitelmänä tai vertaisarvioituna artikkelina jossakin alan arvostetussa julkaisussa. Pelkkä havainto ja sen raportointi eivät siis vielä ole sellaista tieteellistä yleistettävää tietoa, joka on tieteen päämääränä.

*Tutkimus* on yleisnimi systemaattiselle informaation tuottamiselle tai hankinnalle. Monissa yrityksissä tehdään tutkimusta, ja esimerkiksi tutkiva yksityisetsivä tai journalistikin tekevät tutkimusta. Tieteellinen tutkimus eroaa tällaisesta muusta tutkimuksesta, sillä sitä ohjaavat erityiset tieteen pelisäännöt. Tiede on toimintana normatiivista, ja tieteellistä metodia voi pitää yhtenä niistä säännöistä, joita tieteellisen tutkijan on noudatettava. Muita tieteen pelisääntöjä ovat tiedon perusteltavuus, totuudellisuus, objektiivisuus ja julkisuus.

Tieteen pelisääntöihin kuuluvat:

- tieteellinen metodi
- julkisuus
- perusteltavuus
- totuudellisuus
- objektiivisuus

Tieteen normatiivisuuden voi yleensä ymmärtää kahdella tavalla. Esimerkiksi Hans Reichenbach (1938) erotti toisistaan tieteen normatiivisuuteen eli tieteen pelisääntöihin liittyvät kysymykset: Millä keinoin tietoa saavutetaan? Miten tietoa oikeutetaan? Kun tieteen pelisäännöt ymmärretään keinona päästä tavoitteeseen, sääntöjä voi verrata kakkureseptiin: reseptissä annetaan ohjeet siihen, mitä pitää tehdä ja missä järjestyksessä, jotta kakusta tulee sellainen kuin on tavoitteena. Toisaalta tieteen pelisäännöt voidaan ymmärtää tiedon arviointikriteereinä, ja ne kertovat, miten tietoa oikeutetaan. Tuolloin säännöt toimivat samalla demarkaatiokriteereinä, eli ne erottavat tieteen ja pseudotieteen toisistaan. Pelisääntöjen avulla voimme arvioida meille esitettyä tietoesitystä sen perusteella, onko tiedon hankinnassa noudatettu tieteelle tunnusomaisia piirteitä. Tällöin tieteen pelisääntöjä voi verrata rikostutkintaan: jotta rikostutkimuksen tulos on oikeudessa hyväksyttävä, tutkijoiden pitää noudattaa hyväksyttyjä ohjeita ja rutiineja.

Kitcher (1983) havainnollistaa tieteen pelisääntöjen roolia vertaamalla tilannetta siihen, että esitetään kaksi kilpailevaa väitettä todellisuudesta, joista toinen perustuu jumalalliseen ilmoitukseen ja toinen monimutkaiseen matemaattiseen todistukseen. Jumalallisen ilmoituksen maailman luomisesta saanut kreationisti vetoaa ilmestystietoon. Kvanttimekaniikan tutkija esittää väitteitä näkymättömistä partikkeleista vedoten monimutkaiseen matemaattiseen kaavaan. Äkkiseltään voi näyttää siltä, että molemmat väitteet perustuvat intuitioon ja edustavat näkymättömiä asioita ja että kummankin väitteen todeksi hyväksyminen perustuisi pelkkään auktoriteettiuskoon. Kvanttimekaniikan laskelmat aukeavat kuitenkin alaan perehtyneelle samoin kuin esimerkiksi kognitiivisen kieliopin piktografiset kuviot kognitiivisen kielentutkimuksen harjoittajalle. Tieteessä väitteitä voi siis tarkastella kuka tahansa alan teorioihin ja metakieleen perehtynyt. Tieteellisiä väitteitä voi koetella uuden näytön tai uusien tulkintojen valossa. Tieteessä luotammekin viime kädessä siihen, että kullakin alalla asiantuntijat kehittävät tieteen pelisääntöjä kriittisesti ja että väitteitä testataan ja koetellaan niille kuuluvalla tavalla.

Tieteellisen väitteen tai teorian hyveitä:

- Väitteet eivät ole ristiriidassa kokemuksen kanssa.
- Väitteet ovat yhtäpitäviä kokemuksen kanssa.
- Väitteet ennustavat ilmiöitä.
- Väitteet ovat teoreettisesti päteviä ja tuottavat uutta teoretisointia.
- Väitteet selittävät yllättäviä tai uusia ilmiöitä.

Väitteiden testaaminen on joskus kielentutkimuksessa 1900-luvulla ymmärretty kapeasti vain luonnontieteellisen koeasetelman kehyksessä. Nykyään se ymmärretään laajemmin tutkimuksena, jossa väitteiden vastaavuutta, ennustavuutta ja selittävyyttä tarkastellaan uusien aineistojen, tulkintojen tai ilmiöiden valossa. Esimerkiksi kielentutkijan hypoteesia Z voidaan testata uusissa aineistoissa tai murteissa tai tutkia diakronisesta perspektiivistä. Tällainen väitteiden testaaminen tieteessä tuottaa ajan kuluessa vivahteikkaamman kuvan sekä selityshypoteesista Z että ilmiöstä Y. Huomataan esimerkiksi, että *eiku*-korjaus toimii joissain murteissa tai puhetavoissa yhdellä tavalla ja toisissa toisella tai että *eiku*-partikkelilla on erilaisia indeksikaalisia vivahteita, jotka vaikuttavat vuorovaikutuksen kulkuun. Hypoteesia Z ei siis pyritä falsifioimaan 1930-luvun popperilaisessa joko–tai-hengessä, vaan sitä tutkitaan monesta eri näkökulmasta ja kehitetään eteenpäin niiltä osin, kuin se toteuttaa tieteellisen väitteen hyveitä.

Thagard (1978) esittää, että tieteen normatiivisuus nousee viime kädessä yhteisön tarpeesta vaalia tieteellistä tietoa ja sen statusta tavoiteltavana tiedon lajina. Kun viemme lapsemme lääkäriin, kielenopetukseen tai logopedin luo, emme halua, että alan ammattilainen soveltaa päivän horoskooppia vaan viimeisintä tutkimustietoa. Thagardin (mt.) mukaan käsitys siitä, mikä on viimeisintä tutkimustietoa, on tieteellisen yhteisön arvio siitä, miten tutkimustietoa tulee kulloinkin arvioida, koetella, keskusteluttaa ja soveltaa. Horoskooppikin ennustaa, selittää ja perustuu erilaisiin teorioihin. Tieteellisellä tiedolla on kuitenkin yhteiskunnan silmissä lisäarvoa muihin tiedon lähteisiin verrattuna. Pidämme tieteellistä tietoa yhteishyvän lähteenä, ja sen tuottaminen yhteisön tarpeisiin perustuu eräänlaiseen yhteiskunnalliseen sopimukseen, johon

tiedeyhteisö on sitoutunut. Yliopiston myöntämät tutkinnot välittävät muulle yhteisölle tiedon siitä, kenellä on riittävä tieteellinen osaaminen jollakin alalla.

Ilkka Niiniluoto (2001) on erottanut yhteishyvään pyrkivän tieteen ja salen. Termi *sale* on sanojen *salaisuus* ja *vale* yhdistelmä. Sale-tutkimus ei Niiniluodon mukaan täytä julkisuuden kriteerejä, vaan se sitoutuu yleensä kaupallisiin tai valtiollisiin intresseihin ja salaamiseen. Tiedeyhteisö ei siis voi arvioida, onko tutkimusprosessi luotettava ja saavutettu informaatio tosi. Tutkimuksen tuloksia ei myöskään julkaista yhteisön kollektiivisena omaisuutena, vaan tavoitteena on harvojen hyöty. Niiniluoto argumentoi, että sale ja tiede, tutkimus ja tieteellinen metodi, informaatio ja tieteellinen tieto on erotettava toisistaan. Sale on tilaustutkimusta, jonka arvostelukriteerit on yleensä ilmaistu huonosti ja ne jättävät paljon tilaa henkilökohtaisille mieltymyksille. Tieteellisen tiedon luonteeseen kuuluu objektiivisuus, joka on henkilökohtaisten mieltymysten vastakohta: tutkimuksen tulokset eivät ole pelkästään tutkijan tai tutkimuksen tilaajan subjektiivisista ominaisuuksista kiinni. Juuri julkinen tieteellinen prosessi varmistaa tutkimuksen toistettavuuden ja väitteiden intersubjektiivisuuden. Kun tutkimus voidaan periaatteessa toistaa tai kun voidaan ymmärtää eksplikoitua argumentointiketjua seuraamalla, miten väitteeseen on päästy, voidaan verifioida, että väitteet ovat reaalitodellisuuden ominaisuuksista riippuvaisia eli mahdollisimman objektiivisia väitteitä todellisuudesta.

Tiede perustuu Robert Mertonin (1973 [1942]) kiteyttämiin ihanteisiin, jotka ovat Niiniluodon (2001) mukaan seuraavat:

– Universalismi: väitteiden totuusarvon arviointiin eivät saa vaikuttaa väitteen esittäjän henkilökohtaiset ominaisuudet, kuten rotu, kansallisuus, yhteiskuntaluokka tai uskonto. Kriteerien on oltava persoonasta riippumattomia ja kollektiivisesti hyväksyttyjä.
– Kommunismi: tieteelliset keksinnöt ovat yhteistä omaisuutta, koska ne ovat sosiaalisen yhteistyön tuloksia.

– Pyyteettömyys: tutkijan oma ura tai arvovalta ei saa vaikuttaa tieteellisen tiedon etsintään ja erittelyyn.
– Järjestelmällinen epäily: tieteen tuloksia tulee tarkastella kriittisesti sekä todellisuutta että esitystavan loogisuutta vasten.

Ketosen (1976) mukaan keskeinen osa tiedettä on se, että tieteellinen väite – silloinkaan kun se on saavutettu matemaattisella, filosofisella tai kielitieteellisellä intuitiolla – ei ole väite tutkijan subjektiivisista ominaisuuksista vaan jostakin niistä riippumattomasta, erillisestä tai yhteisesti jaetusta ilmiöstä. Esimerkiksi Croftin (2001) tietojäsennys substantiivista ei pyri kertomaan tutkijasta itsestään vaan tietystä ilmiöstä. Tietojäsennysten, väitteiden ja tieteellisten konstruktioiden tueksi esitetään julkisia analyyseja ja perusteluja, joilla on tieteellisen argumentaation rakenne, vaikka itse päätelmä on ehkä alkujaan ollut tutkijan oman intuitiivisen työn tulosta.

Tieteellisen metodin ja argumentaation julkisuus mahdollistaa tieteen itseään korjaavan luonteen. Tiedeyhteisö suhtautuu kriittisesti saatuihin tuloksiin; se toistaa tutkimusteemoja ja kehittää tutkimuksia edelleen. Toisin kuin esimerkiksi uskonnolliset tai ideologiset väitteet tieteellinen väite ei perustu auktoriteettiin vaan siihen, että väitteet asetetaan alttiiksi jatkuvalle koettelulle, ja ne hylätään, mikäli reaalitodellisuus on väitteiden kanssa ristiriidassa. *Reaalitodellisuus* on termi, jolla viitataan tietoesityksen eli tieteellisen taustateorian mallintamaan ulkopuoliseen maailmaan (ks. Johdannon luku 2.4. tk.). Termin *reaalitodellisuus* avulla voidaan siten erottaa toisistaan teorian mallintama maailma ja se maailma (reaalitodellisuus), josta se esittää väitteen. Esimerkiksi väite ”suomen kielen transitiivilauseessa voi olla partitiivisubjekti” (esim. ***Monia opiskelijoita*** *tekee kahta tutkintoa*) on tosi ainoastaan siinä tapauksessa, että reaalimaailmassa vallitsee lauseen ilmaisema tila eli että kyseisenlaisia transitiivilausetyypin esiintymiä löytyy.

Teoreettiset mallit käyttävät hyväkseen metakäsitteitä, kuten *partitiivi*, *subjekti* tai *transitiivinen*, joiden avulla luokitellaan ilmiöitä. Metatermit määritellään koherentisti jossakin taustateoriassa, esimerkiksi jossakin kielenkuvausmallissa, ja niillä on sekä aineistoa että teoriaa

systematisoiva funktio. Kun tutkija esimerkiksi esittää, että partitiivisubjektia käytetään transitiivirakenteissa, ja osoittaa aineistoesimerkein luokitusten pitävän paikkansa, kyse on väitteen oikeuttamisesta.

Kielentutkimuksessa kieliopilliset luokitukset ja kieliesimerkit muodostavat keskeisen osan tieteellistä argumentaatiota. Tyypillistä on osoittaa esimerkiksi erilaisten minimipariesimerkkien avulla, että jokin luokittava metatermi toimii tai luokitusta koskeva yleistettävä väite pitää paikkansa. Esimerkit voivat olla peräisin luonnollisesta aineistosta, tai kielentutkija voi kielitajunsa avulla esittää kohdennettuja täsmäesimerkkejä. Säännönmukaisuuksien osoittaminen aineistosta oikeuttaa kielitieteellisessä argumentaatiossa päätelmän eli yleistettävän kuvauslauseen kielestä. Tieteen päämääränä ovatkin juuri yleistettävät lauseet (esim. Jos Y kuuluu X:ään, niin sillä on ominaisuus Z). Tieteellinen teoria voidaan ymmärtää yleistettävien lauseiden tai yleispätevien kuvauksien joukkona, jonka perusteella voidaan ennustaa, miten ilmiöt käyttäytyvät ja millaisia suhteita niillä on keskenään.

Entä mitä ovat yleistettävät lauseet tai yleispätevät kuvaukset? Vastaus kysymykseen, ovatko yleistettävät lauseet tai kielentutkimuksessa käytettyjen termien referentit universaaleja, luonnollisia kategorioita, lakeja vai vain instrumentteja, joita käytetään aineiston systematisaatiossa (ks. Johdannon luku 2.4. tk.), jakaa kielentutkimuksen teoriat erilaisiin luokkiin. **Käsiterealistit** olettavat, että kieltä kuvaavat metakäsitteet ovat jollakin tapaa itsenäisesti olemassa universaaleina. **Nativistit** perustelevat universaalien metafysiikkaa biologisin argumentein, ja **empiristit** yleensä kieltävät *laki*-käsitteen tyystin ja puhuvat reaalimaailman lainomaisista (noomisista) ilmiöistä. **Nominalistit** ja **instrumentalistit** vuorostaan esittävät, että todellisuus on ihmisen konstruoimaa ja metatermit ovat vain käteviä työkaluja. Heidän mukaansa termit, kuten *partitiivi* tai *transitiivinen*, eivät siis viittaa universaaleihin kielen kategorioihin, jotka ovat jollakin tapaa olemassa ihmisestä riippumatta, vaan ne ovat kielellisiä välineitä kielentutkijan työkalupakissa.[1]

Riippumatta siitä, mikä koulukunnan käsitys on metateoreettisten käsitteiden ontologiasta, yleensä ollaan yhtä mieltä siitä, että tieteellisten termien määritelmiä korjataan tutkimuksen kuluessa. Itseään korjaavan luonteensa vuoksi tiedettä pidetään usein myös edistyvänä. Termille

*tieteellinen edistys* annetaan yleensä kaksi eri tulkintaa. **Veristisissä teorioissa** eli teorioissa, joissa tieteen päämääränä katsotaan olevan totuus, tieteen uskotaan hiljalleen lähestyvän totuutta. Esimerkiksi **kriittisen tieteellisen realismin** mukaan tiede ei ehkä koskaan saavuta totuutta mutta siinä lausutut väitteet mukautuvat tieteellisessä prosessissa totuuden kaltaisiksi (Niiniluoto 1980b). **Instrumentalistisissa teorioissa** vuorostaan katsotaan tieteen edistyvän, kun tieteellisten väitteiden käyttöala laajenee tai paranee. Käytettävyys ilmaisee todellisuuden ja kielen välisen mielekkään suhteen. Kun jonkin tieteellisen väitteen tuloksena on kantava silta, toimiva koje tai vaikkapa vieraan kielen opetukseen sopiva kielioppi, on väite käyttökelpoinen ja se voidaan hyväksyä tieteelliseksi.

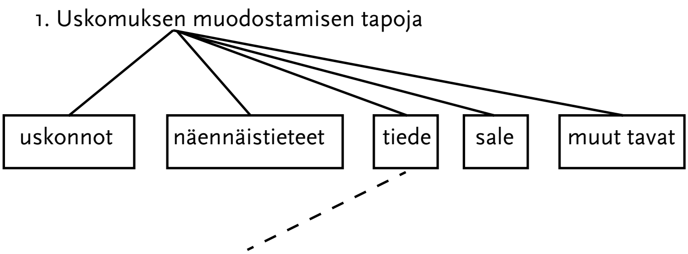
Uskomuksen muodostamisen eri muodot, kuten reflektio, abduktio ja intuitio, kuuluvat kaikilla tieteenaloilla tutkijan työkaluihin. Ero pelkän uskomuksen ja tieteellisen tiedon välillä liittyy siihen intersubjektiiviseen julkiseen prosessiin, joka ohjaa tiedettä. Näennäistieteessä 1) uskomukseen johtaneita keinoja ja askelia ei yleensä selvennetä julkisesti, 2) selitys ei noudata tieteellistä argumentaatiota, 3) tulokset eivät ole toistettavissa, eli ei voi koettaa, pääseekö samoihin tuloksiin uudella aineistolla, 4) tulokset ovat uskomuksen luojan subjektiivisista ominaisuuksista riippuvaisia tai 5) väitteitä ei voida testata uusilla kokeilla.

Näennäistieteen ja tieteen rajaa pohdittaessa on mielekästä tarkastella tieteellisyyden kriteerejä alakohtaisesti. Tällöin puhutaan siitä, missä määrin ala tai tutkimussuuntaus on **idiografinen** tai **nomoteettinen**. Esimerkiksi historiantutkimusta pidetään idiografisena tieteenä, joka tutkii ainutkertaisten historiallisten tilanteiden luonnetta ja jossa tutkimus rajoittuu tapahtumahistoriaan. Esimerkiksi kielitieteen historian tai yksittäisten kielenmuistomerkkien (vanhat tekstit yms., ks. Näkökulmia filologiseen tutkimukseen tk.) tutkimus kartoittaa tilanne- ja tapahtumahistoriaa. Aineistoa ei voi vaihtaa eikä testata uudestaan, vaikka samaa aineistoa voikin tulkita uudelleen. Tutkimuksen tieteellisyys perustuu päättelyprosessien ja lähteiden julkiseen tarkasteluun ja tieteellisen argumentaation rakenteeseen. Psykolingvistiikkaa tai (labovilaista) sosiolingvistiikkaa voidaan taas pitää nomoteettisina aloina, joiden tavoite on löytää tilastollisesti merkittäviä lainomaisuuksia ilmiöistä (ks. Psyko- ja neurolingvistiikka tk.).

Asetelma nomoteettisen ja idiografisen tutkimuksen välillä voi kuitenkin vaihdella jopa yhden ja saman tutkimuksen sisällä. Kielentutkijat ovat yleensä kiinnostuneita sekä ainutlaatuisista ilmiöistä että yleisemmistä kielen kehitys- ja muutosilmiöistä. Toisinaan kielentutkimus keskittyy juuri idiografisen ja nomoteettisen suhteeseen: kielentutkijat kiinnittävät huomiota konkreettisessa, ainutlaatuisessa puhetilanteessa esiintyvään yleiseen seikkaan. Yleinen tarkoittaa teoriaa, jossa on yleispäteviä kuvauksia. Kun kielentutkimuksessa halutaan paljastaa ilmiöiden yleisiä ominaisuuksia, täytyy sitä varten olla käytössä metodologinen taustateoria, joka tarjoaa välineitä yksittäisen ilmiön tai havainnon luokitteluun ja vertailuun.

Tieteen kriteerit ovat näennäistieteen vastakohtia: 1) uskomukseen johtaneita keinoja ja askelia selvennetään julkisesti, 2) selitykset noudattavat tieteellistä argumentaatiota, 3) tulokset ovat toistettavissa, eli voi koettaa, pääseekö samoihin tuloksiin uudella aineistolla, 4) tulokset eivät ole uskomuksen luojan subjektiivisista ominaisuuksista riippuvaisia ja 5) väitteitä voidaan testata uusilla kokeilla tai muuten arvioida. Kaikki kohdat 1–5 sisältävät intersubjektiivisen osatekijän, jonka voidaan siten ajatella kuuluvan tieteellisen metodin määritelmään: metodi on tieteellinen, kun se on tieteellisen yhteisön jakama ja johtaa yhteishyvään eli yhteiskunnan jäsenten tasapuolisesti ulottuvilla olevaan tieteelliseen tietoon.

Edellä on käsitelty tieteen normatiivista luonnetta ja niitä erityispiirteitä, jotka erottavat tieteen muista uskomuksen muodostamisen lajeista. Kuviosta 8 huomataan, että kollektiivisia uskomuksen muodostamisen tapoja on useita, joista tiede on vain yksi.




Kuvio 8. Uskomuksen muodostamisen tapoja.

Seuraavaksi pohditaan tarkemmin sitä, millä tavalla tieteellistä uskomuksen muodostamisen tapaa voidaan paremmin ymmärtää tietoteorian kautta. Tietoteoria vastaa muun muassa seuraaviin kysymyksiin: Miten tietoa saavutetaan? Mistä tietää, että uskomus on tietoa? Voiko tieto olla tietoa ilman, että kukaan osaa sitä perustella?

## 2.3. Tietoteoreettiset ehdot

Tieto-oppi (tietoteoria, epistemologia) on ala, joka tutkii tietämistä ja tietoa. Useimpien tietoteoreettisten[2] käsitysten mukaan tieteellinen tieto ei ole ainoa tietämisen muoto. Tieto ja tieteellinen tieto eivät siis ole synonyymejä. Matti voi esimerkiksi tietää, että hänen nimensä on Matti, että koiraa kutsutaan suomen kielessä *koiraksi* tai että Matilla itsellään on kaksi kättä. Tällainen subjektiivinen yksittäistieto ei kuitenkaan ole tieteellistä, koska tiedon subjekti ei ole tässä tapauksessa kollektiivi eli tiedeyhteisö ja tiedon kohteena olevat propositiot eivät ole käyneet läpi tieteelliselle väitteelle kuuluvaa julkista prosessia eli julkaisua, väittelyä, koettelua ja vakiintumista.

Tieteellisellä tiedolla on myös muita erityisominaisuuksia. Tieteessä tavoitteena ovat yleispätevät kuvaukset: niillä katsotaan olevan ennustavaa arvoa, jonka avulla voidaan tehdä deduktiivisia päätelmiä. Jos esimerkiksi tiedämme, millaisia yleisiä ominaisuuksia ihmisillä on, voimme tämän tiedon perusteella johtaa yksittäistapauksia koskevia päätelmiä ja tällä tavoin myös koetella teoriaa (esim. ”Matti on ihminen, joten hänellä on lajimääritelmän mukaan edellytykset siihen, että hänellä on kaksi kättä”). Tieteellinen teoria voidaankin määritellä juuri yleispätevien kuvausten joukoksi, joka antaa koherentin kokonaiskuvan jostakin ilmiöstä. Toisin sanoen Matin henkilökohtainen tieto siitä, millä suomen kielen sanalla koiraa nimitetään, ei vielä yksinään kiinnosta tiedettä, mutta tämä tieto esimerkiksi osana leksikaalista semantiikkaa voi johtaa mielenkiintoisiin tutkimuskysymyksiin, joiden vastauksissa on kieltä koskevia yleispätevyyksiä.

Ollakseen tietoa yleistettävällä lauseella tulee olla totuusarvo. Totuutta koskevat teoriat perustuvat tietoteoriaan. Platon loi nykytieteen

perustana olevan klassisen tietoteorian määrittelemällä tiedon hyvin perustelluksi todeksi uskomukseksi teoksessaan *Theaitetos*. Tietämisen loogisen rakenteen voi esittää seuraavasti:

A tietää, että P, jos ja vain jos
- A uskoo, että P.
- A:lla on pätevät perustelut uskoa, että P.
- P on tosi.

Tiedon määritelmäksi ei siis riitä se, että A uskoo, että hän on voittanut lotossa, ja että uskomus toteutuu eli A todella on voittanut lotossa. Mikäli A:lla ei ole perusteita uskomukselleen, kyseessä ei ollut tieto vaan pelkkä uskomus. A ei tiennyt voittavansa, koska hänellä ei olisi ollut päteviä perusteita uskolleen. Tieto on siis relaatio subjektin kognitiivisen tietämisen tilan (uskomus), maailmassa vallitsevan asiantilan (tiedon kohde) ja uskomusta puoltavien argumenttien (tiedon sisältö) välillä.

Maallikon A sujuva kielitaito ei sekään ole tieteellistä tietoa, vaikka kaikki A:n uskomukset kielestä L olisivatkin oikeita eli tosia. Tieteellisen tiedon perustelut ovat julkisia, ja tiedon subjekti on tiedeyhteisö. Tämä tarkoittaa sitä, että tutkimusjulkaisun päämäärä on muuttaa tiedeyhteisön kognitiivinen asenne epäuskosta uskoon esittämällä vakuuttavasti perustelut sille, miksi P on tosi. Tieteellinen argumentaatio perustuu siis vakuuttavien argumenttien esittämiselle.

Tieteellisen argumentaation tulee olla loogista ja koherenttia ja siinä olevien merkityssisältöjen perustua yhteen yhtenäiseen käsitejärjestelmään. Oleellinen osa tieteellistä argumentaatiota on tieteellinen käsitteenmuodostus. Teoreettisen termin referentti (eli tarkoite tai kohde) on teoreettinen konstruktio, jonka määritelmä on osa jotakin tieteellistä teoriaa. Termeillä *genetiivi* ja *kvantti* on olemassa luonnollisessa puheessa yleensä maallikon ymmärtämä yleismerkitys, mutta tieteenharjoittaja tuntee alansa koulukuntakiistat ja tieteellisen termin aseman käsitejärjestelmän kokonaisuudessa ja siksi harvoin pitää teoreettisten termien merkityksiä lukkoon lyötyinä. Teoreettisten termien määritelmä vaihtelee myös eri teorioissa ja muuttuu tutkimuksen

myötä. Harvinaista ei ole sekään, että tutkimuksen kuluessa ilmenee sellaisia yllättäviä seikkoja, jotka muuttavat niin tutkijan omaa kuin tiedeyhteisönkin maailmankäsitystä ja näin myös termin sisältöä.

Tieteessä erilaiset teoreettiset ja operationaaliset termit muodostavat havaintokäsitteistön. Esimerkiksi erilaiset fonetiikan ja fonologian luokitukset kategorisoivat äänihavaintoa. Varsinaisen havaintokeinon muodostaa itse tieteellinen metodi, ja taustateoria taas kertoo, millä tavalla ilmiöt liittyvät toisiinsa (Putnam 1977). Usein viittausten kohteena olevaa reaalimaailmaa edustaa jokin aineisto, johon viitataan. Sellaisen väitteen kuin ”partikkelilla *eiku* reagoidaan välittömästi korjausta vaativaan seikkaan” tueksi voidaan tieteellisen argumentaation kuluessa vedota aineistoesimerkkeihin, joissa tämä havaittu seikka ilmenee (ks. Keskustelunanalyysi ja vuorovaikutuslingvistiikka tk.). Aineiston hankintatapa on kullekin kielentutkimuksen haaralle ominainen ja täyttää sen teorioiden asettamat kriteerit. Tutkittaessa ilmiöitä, joissa esiintyy suurta vaihtelua ja variaatiota, kuten keskusteluja, äänitetään yleensä kymmeniä tunteja erilaisia keskustelutapauksia, kun taas ääntämistä koskevia rakenteita eli fonetiikkaa tutkittaessa pienempikin aineisto riittää.

Tieteellisen argumentaation ja analyysin vakuuttavuus sisältyy tutkijan puhetekoon, joka johtaa ihanteellisessa tapauksessa uudistavaan vaikutukseen tiedeyhteisössä: epäusko muuttuu uskoksi siten, että tiedeyhteisö alkaa tietää jotakin.

Tiedeyhteisö uskoo, että P.

Tiedeyhteisöllä on pätevät perustelut uskoa, että P.

P on tosi.

Tieteessä tutkimusjulkaisu koostuu väitteestä ja sen perusteluista. Perustelujen tavoitteena on vakuuttaa tiedeyhteisö väitteen todenperäisyydestä. Perusteluja on karkeasti katsottuna kahdenlaisia, ja niiden mukaan voidaan tieteenalat jakaa karkeasti formaalitieteisiin ja reaalitieteisiin.[3] Tätä jakoa käyttäen voidaan havainnollistaa eroja päätelmien tekemisen ja todistamisen menetelmän välillä.

**Formaalitieteitä** ovat esimerkiksi logiikka ja matematiikka. Propositiologiikan ja matematiikan aksioomajärjestelmä tarjoaa todistamisen menetelmän eli menetelmän, joka todistaa tai tuottaa oikeita (tosia) propositiologiikan tai aritmetiikan teoreemoja. Kuka tahansa, joka tuntee merkkien 2 + 2 = 4 merkityksen, vakuuttuu päätelmän todenmukaisuudesta, koska aksioomajärjestelmään perustuva argumentaatio vakuuttaa järkemme. Järkemme vakuuttuu tahtomattamme, sillä aksioomajärjestelmä perustuu Aristoteleen (1994) mukaan loogiseen välttämättömyyteen: teoria ilmaisee asiantiloja, jotka eivät voisi olla muuten kuin ovat.

Aksioomajärjestelmistä johdettuja tosia lauseita on pidetty universaaleina, analyyttisina ja välttämättöminä. Perustelut todistavat lauseet. Toisin sanoen aksioomajärjestelmistä johdetut lauseet eivät ole esimerkiksi kieliyhteisön, kulttuurin tai aikakauden normi- ja merkitysjärjestelmistä riippuvaisia: 2 + 2 ei kehittynyt kolmesta neljäksi, eikä tuloksen voi odottaa muuttuvan ajan kuluessa viideksi. Jos joskus löydämme toiselta planeetalta elämää, on varsin todennäköistä, että universumin toisellakin puolen määrien yhdistäminen noudattaa periaatetta 2 + 2 = 4.

Kielentutkimus kuuluu **reaalitieteisiin** siitä huolimatta, että se on usein teoreettista ja käyttää toisinaan aksiomaattisia tai aksiomaattistyyppisiä järjestelmiä kielen analyysissa. Reaalitieteisiin voidaan lukea kaikki sellaiset tieteet, joiden tutkimuskohde kuuluu reaalimaailmaan ja joissa väitteen voi kumota tai väitettä voi vahvistaa reaalimaailmaa koskevalla aineistolla.

Joissain reaalitieteissä, kuten kielentutkimuksessa, on mielekästä tehdä ero normiproposition ja normilauseen välillä. **Normipropositio** on väite todellisuudesta (esim. ”*talossa*-sana on suomen kirjoitetun yleiskielen *talo*-sanan yksikön inessiivi”), kun taas **normilause** on ilmaus korrektista tai oikeaksi oletetusta ilmiöstä (esim. ”*talossa* on korrektia suomea”). Myös esimerkiksi oikeustiede tutkii yleensä normipropositioita (mitä laki tosiasiassa on) ja oikeusdogmatiikka normilauseita (mitä lakien tulisi olla). Normilauseet johdetaan yleensä järjestelmän sisäisistä suhteista, kun taas normipropositiot vuorostaan viittaavat johonkin maailmassa vallitsevaan tilaan. (Oikeustieteestä ks. esim. Kaisto 2005.)

Kielijärjestelmä itsessään on loogisesti kontingentti eli ei-välttämätön. Kontingenttia on esimerkiksi se, että vaikka on totta, että Sauli Niinistö

oli vuonna 2017 Suomen tasavallan presidentti, silti olisi voinut olla myös toisin. Voidaan esittää joukko tapahtumia, jotka uskottavasti olisivat voineet muuttaa historian kulkua. Ei siis ole välttämätöntä, että Niinistö olisi ollut tasavallan presidentti vuonna 2017, vaikka väite on tosi. Kontingentit väitteet ovat kumoutuvia ja korjautuvia: osa historian, lääketieteen ja kielentutkimuksen väitteistä korvataan uusilla väitteillä, tai ne kehittyvät tarkemmiksi tieteen edistyessä. Usein väitteen ala eli ekstensio muuttuu myös reaalitodellisuudessa: vuoden 2067 presidenttiä emme tiedä varmuudella sen perusteella, että tunnemme nykytilan. Suomen kieli järjestelmänä on pitkän kehityksen tulos. Emme varmasti voi siksi myöskään sanoa, miltä suomen kieli näyttää sadan vuoden kuluttua.

Kielentutkimuksen ja yleensäkin ihmistieteellisen tiedon erityisluonteeseen kuuluu myös, että tieteellinen tieto voi muuttaa tutkimuskohdettaan. Oppikirja Newtonin painovoimakäsityksestä ei voi vaikuttaa painovoimaan, mutta kielioppi voi vaikuttaa kieliyhteisöön ja pystyy vaikuttamaan sen jäsenten käytökseen (Määttä 2000). Ihmisen kielellinen käytös ei siten pelkästään ole kontingenttia ja ajassa muuttuvaa, vaan se voi myös muuttua tieteellisin keinoin tehdyn kuvauksen vaikutuksesta (mt.). Usean kielen tutkimuksessa voidaankin määrittää aika ennen ensimmäistä (koulu)kielioppia ja sen jälkeinen aika. Kieliyhteisön käsityksillä (standardi)kielestä on taipumus vaikuttaa kielen kehitykseen.

Kieli on siis muutosherkkyydessään hyvin erilainen kuin formaalitieteen aksioomajärjestelmä, joka on muuttumaton, ja juuri siksi se on perustelujen muotokieli. Reaalitieteissä vuorostaan käytetään luonnollista kieltä tieteellisessä argumentaatiossa ja erilaisia tieteellisiä metodeja, joiden tukena hyödynnetään toisinaan matemaattis-tilastollisia ja loogisia välineitä. Reaalitieteissä empiirisellä näytöllä eli evidenssillä on argumentaatiossa yleensä keskeinen rooli. Käsitettä *empiirinen evidenssi* ei tule tässä kontekstissa ymmärtää liian kapeasti esimerkiksi tiukasti aikaan ja paikkaan sidottuna havaintometodein saatuna aineistona vaan erityisesti reaalimaailmaan viittaavana materiaalina, jota vasten väitteitä ja teorioita pitkällä aikavälillä verrataan.

Reaalitieteissä on eri aloilla käytössä erilaisia menetelmiä, joiden avulla aineistoja a) kerätään, b) systematisoidaan ja kategorisoidaan

sekä c) tulkitaan. Eräissä metodeissa aineiston keruu ja tulkinta liittyvät tiiviisti toisiinsa. Menetelmä on prosessi, jonka aikana kategorisointi, mittaus ja joiltain osin myös tulkinta tapahtuvat samanaikaisesti. Reaalitieteellisissä julkaisuissa on siksi oleellista avata tutkimusasetelma ja prosessin kulku mahdollisimman tarkasti.

Tieteellisessä argumentaatiossa menetelmän kuvaus muodostaa osan väitteen perustelua ja selittää sen, miten väitteeseen on päädytty ja miksi tiedeyhteisön tulisi hyväksyä väite todeksi. Kielitajun varassa operoivan kielentutkimuksen voi ajatella kielentutkimuksessa muodostavan tieteellisessä argumentaatiossa poikkeuksen, sillä se ei näyttäisi aina selventävän tarkasti tutkimusaskeliaan. Implisiittisen kielitiedon systematisaatio on kuitenkin kielentutkimuksessa välttämätön aktiviteetti, johon perustuu moni muu alan tutkimustekniikka (ks. Kielitaju kielentutkijan työkaluna tk.).

Jako formaalitieteisiin ja reaalitieteisiin on karkea, ja tällaisen suurpiirteisen kategorisoimisen motivaatio ja konteksti on yleensä ollut tietoteoreettisten käsitteiden tai tieteenfilosofisten perusongelmien selventäminen. Tieteenfilosofia itse on käsiteanalyysia (ks. Käsiteanalyysi kielentutkimuksessa tk.), eikä se siten sovi ongelmitta kumpaankaan kategoriaan. Käsitteiden analyysissa eritellään ja selkeytetään tunnettujen termien ja käsitteiden merkityksiä ja keskinäisiä suhteita. Tieteenfilosofiassa käytetään usein tukena myös formaalis-loogisia välineitä. Tieteenfilosofia ei kuitenkaan tästä huolimatta ole formaalitiede – kuten ei kielentutkimuskaan –, jossa päädyttäisiin universaaleihin ja ajattomiin totuuksiin tieteellisten käsitteiden merkityksestä.

1900-luvulla käsitys tieteestä oli nykyistä kapeampi ja tieteen kenttä jaettiin usein dikotomisesti luonnontieteisiin ja formaaleihin tieteisiin unohtaen useimmiten tyystin humanististen tieteiden erikoisluonne. Tässä kahtiajaossa pyrittiin luokittamaan kaikki tieteenalat. Siksi esimerkiksi kysymys siitä, mikä on kielitieteen perimmäinen luonne, puhutti joitakin tutkijoita Suomessa 1970-luvulla: onko kielitiede formaalia tiedettä vai luonnontiedettä?[4] Chomsky ja autonomisen kielitieteen koulukunnan jäsenet esittivät, että kielitiede perustaisi metodologiansa rationalistiseen premissiin eli käsitykseen siitä, että kielentutkimuksen kuuluu kohdistua sisäsyntyisiin universaaleihin rakenteisiin, joihin

tutkijalla on intuitiivinen pääsy. Chomskyn käsityksen mukaan kielitiede oli luonnontiedettä mutta samalla formaalia analyysia. Autonomista kielitiedettä vastustavat taas väittivät päinvastaista.

1970-luvun suomalaista keskustelua kielitieteen luonteesta leimaa ainakin kaksi käsitteellistä vääristymää, jotka johtivat usealla taholla erheellisiin päätelmiin. Ensiksi kielitiede samastettiin yksinomaan autonomiseen kielitieteeseen (ks. esim. Lyas 1971). Empiiristä evidenssiä käyttäviä kielentutkimuksen aloja harvoin sisällytettiin keskusteluun. Samaan aikaan luonnontieteet, empiiriset tieteet ja eksperimentaatio ymmärrettiin usein samana kategoriana. Kategoriavirhe johti päätelmään, että luonnontiede on empiiristä ja eksperimentaalista ja kaikki empiirinen tutkimus vuorostaan välttämättä luonnontiedettä. Käsitys on kuitenkin virheellinen: teoreettinen fysiikka on luonnontiede (mutta ei empiiristä), mutta esimerkiksi geologia ei ole eksperimentaalista (mutta on empiiristä). Taiteen ja kirjallisuuden tutkimus on aineistolähtöistä (empiiristä) mutta harvoin eksperimentaalista – eikä kulttuurintutkimuksena koskaan luonnontiedettä.

Tällaisten aiemmin käytyjen debattien tutkiminen on hedelmällistä sikäli, että ne osoittavat vallitsevien tiede- ja tieteenalakäsitysten voimakkaan vaikutuksen tieteenfilosofisessa keskustelussa. Erilaisten kategoriavirheitten puitteissa keskustelua käytiin *Virittäjä*-aikakauslehdessä 1970-luvulla vieläpä dikotomisesti joko–tai-akselilla: kielitiede nähtiin joko formaalitieteenä tai empiirisenä luonnontieteenä. Ongelmana oli, että kielitiede ei näyttänyt soveltuvan selväpiirteisesti kumpaankaan kategoriaan.

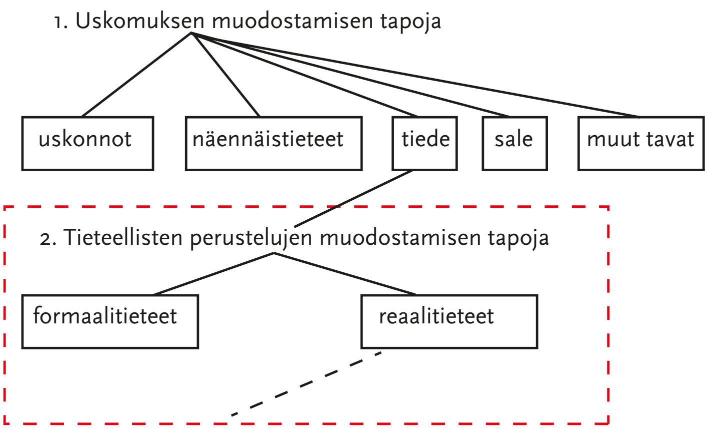
Esimerkiksi Esa Saarisen (1975) mukaan kielioppi ei ole sellainen aksiomaattinen järjestelmä, joka tuottaisi tosia sääntölauseita tai lauseita. Produktiivinen kielioppi tuottaa kieliopillisia lauseita, ei tosia. Kielioppia ei voi osoittaa oikeaksi formaalisten tieteiden ristiriidattomuudella tai täydellisyystodistuksiin verrattavilla yleisillä todistuksilla. Kielitiede olisi siis tässä mielessä empiirinen tiede. Tässä keskustelussa kysymys kielentutkimuksen luonteesta rajoittuu päätelmiin kieliopin mukaisten lauseiden luonteesta. Koko kielitieteen laajaa metodikirjoa tai niitä metodeja ja tutkimusnäkökulmia, joilla jo yksinomaan kielioppia voi tutkia, ei tuolloisessa keskustelussa ole otettu huomioon.

Esa Itkonen (1997, 8–9) antoi joukon a priori -tyyppisiä määritelmiä[5] kielen normeille: ”[j]os niiden [kielen normien] tiedetään olevan tosia, ne eivät ole falsifioitavissa, mikä taas tarkoittaa niiden olevan ei-empiirisiä ja siten tietyssä mielessä apriorisia” (ks. myös Itkonen 1975, 399).[6] Kuten monet muutkin tutkijat Itkonen tavoitteli hyvää määritelmää intuitiiviselle kielitiedolle. Itkonen tunnisti, että kielitieteilijän intuitio on ohittamaton kaikilla kielentutkimuksen aloilla, ei vain generativistisessa metodiikassa. Itkonen käyttää perusteluissaan termiä *falsifikaatio*, joka on Popperin (2002 [1934]) tieteenfilosofiaan kuuluva periaate, joka vaatii, että tieteellisiä teorioita tulee voida testata aineiston valossa. Sen mukaan sellainen teoria, jota ei voida testata, esimerkiksi freudilainen psykoanalyysi tai uskonnollinen dogma, ei ole tieteellinen. Teorian falsifioimattomuus ei kuitenkaan muuta teoriaa tai sen esittämiä väitteitä puhtaan aprioriseksi tiedoksi, kuten tautologiat ”poikamies on naimaton mies” tai 2 + 2 = 4. (Ks. Konstenius 2014, 130–152.)

Tavoittamatta kielitieteen tietoteoreettista perusongelmaa, joka liittyy juuri kielitajun tieto-oppiin, Ketonen (1975), Dahl (1975) ja myöhemmin Niiniluoto (1990) vastasivat Itkoselle, että sosiaaliset normit, kielen sääntölauseet, ovat kontingentisti eli ei-välttämättä tosia, eivät siis *a priori* -totuuksia. Kielioppiesityksen ”totuus” voidaan osoittaa todeksi esimerkiksi käymällä läpi aineistoa, jossa ihmisten tai eri ihmisryhmien todellinen kielenkäyttö toteutuu (Ketonen 1975, 197). Kieli on erilaisten muuttujien yhteisvaikutuksen summa: jos esimerkiksi englannin kielen kehitys alkaisi aivan alusta, se voisi olla joiltain osin toinen ja siten myös kielellisen kehityksen lopputulos, nykyinen englannin kieli, olisi joiltain osin erilainen. (Ks. myös Konstenius 2014, 129–136.)

Edellä kuvattu keskustelu havainnollistaa tapaa, jossa puoli vuosisataa sitten on yläkategorioita käyttämällä pyritty kielitieteen metateoriassa päättelemään jotakin pysyvää kielen ja kielitieteen luonteesta. Modernin tieteenfilosofian mukainen lähestymistapa on tarkastella laaja-alaisesti eri kielitieteen aloja ja kielitieteen metodiikkaa niiden omilla ehdoilla ja päätellä näitä aloja tutkimalla jotakin yleistä kielentutkimuksen luonteesta (Achterhuis 2001). Tutkimuksen suunta on siis alhaalta ylös, konkreettisesta (tutkimustoiminnasta) abstraktiin (käsiteanalyysiin).

Laajamittainen kielitieteen metodiikan tarkastelu osoittaa kielitieteen aseman ja luonteen reaalitieteissä. Reaalitieteiden metodologia alkoi kehittyä varsinaisesti 1600-luvun kuluessa, kun oivallettiin, että keräämällä systemaattisesti erilaisia aineistoja ja rakentamalla tutkimusasetelmia, joilla on looginen rakenne, voitiin perustella pätevästi reaalimaailmaa koskevia väitelauseita. Reaalitieteissä tutkimuksessa käytettyjen aineistojen ja metodien tulee pystyä perustelemaan väite. Väitteiden paikkansa pitävyyttä arvioidaan yleensä erilaisten aineistojen valossa ja eri näkökulmista. Esimerkiksi suomalaisessa kielentutkimuksessa E. N. Setälän astevaihteluteorian perusteluja on selvitelty ja tutkittu laajasti (esim. Korhonen 1986, 129–133). Setälän mukaan astevaihtelu oli kuulunut jo uralilaisen kantakielen morfofonologiseen järjestelmään ja käsittänyt alkuaan koko konsonantiston lisäksi myös vokaaliston. Selvittämällä ilmiön ikää ja levinneisyyttä suomessa ja sukukielissä on Setälän väitteiden perustelut kumottu toistuvasti vastaevidenssin valossa. Juuri kielellisten aineistojen käyttäminen esitettyjen väitteiden ja teorioiden luotettavuuden kriteerinä osoittaa kielentutkimuksen paikan reaalitieteiden joukossa. Siksi kuviossa 9 kielentutkimus on sijoitettu tieteellisten perustelujen pääasiallisen muodostamisen tavan mukaan reaalitieteisiin.




Kuvio 9. Tieteellinen tieto on perusteltu (tosi) uskomus.

## 2.4. Ontologiset ehdot

Edellisessä luvussa käsiteltiin kysymystä, miten kielentutkimuksessa ratkaistaan se, onko väitteen perustelu pätevä, eli mikä on kielentutkimuksen tieto-oppi. Todettiin, että kielentutkimus kuuluu reaalitieteisiin, sillä sen tutkimuskohde on eräs reaalimaailman osa. Reaalitieteille on tyypillistä tarve määrittää tutkimuskohde: Millaiseen ilmiöön maailmassa tutkimus kohdistuu? Miten se havaitaan tai tunnistetaan?

Tässä johdannon luvussa tarkastellaan lähemmin kielen ontologiaa eli sitä, miten kielentutkimuksen kohde määritellään. Jotta missään tutkimuksessa voidaan soveltaa tieteellistä metodia, täytyy ensiksi määritellä tutkimuksen kohde ja kertoa, millä tavoin tästä kohteesta tutkimuksen kuluessa saadaan tietoa. Metodin sovellus reaalitieteissä on siis aina tietynlaisen ontologisen ja tieto-opillisen premissin yhdistelmä. Tieteessä tutkimuskohdetta koskeva määritelmä hyvin harvoin on kuitenkaan vain lauseen mittainen, sillä se on osa laajempaa taustateoriaa. Tällaisen teorian määritelmät sisältävät välttämättä eriasteisia ontologisia sitoumuksia kiinnittäessään tutkimuksen kohteen taustateoriassaan muihin ilmiöihin: ne kertovat, mitä maailmassa on ja miten maailman oliot liittyvät toisiinsa.

Quinen (1948) mukaan tieteellisen teorian käsitejärjestelmä tekee ontologisia sitoumuksia todellisuuden perimmäiseen luonteeseen ja joukkoon yläkäsitteitä. Tieteellinen teoria vastaa onttiseen kysymykseen (engl. *ontic question*) eli kysymykseen siitä, mitä maailmassa on, ja tekee näin eksistenssiväitteitä, joita voidaan puolestaan kutsua *ontologiseksi* tai *onttiseksi listaksi*.

Onttinen kysymys: Mitä on olemassa maailmassa?

$\exists x$ (x on...)

Metafyysinen kysymys: Mistä maailman perimmäinen olemus on tehty?

Quinen (1948) mukaan vastaus onttiseen kysymykseen on listamainen: ”maailmassa on numeroita, olioita, syitä, seurauksia jne.”. Ontologisessa tutkimuksessa käytetään eksistenssikvanttoria ∃ ’on olemassa’. Tieteellisessä teoriassa se kattaa teorian tekemät väitteet ilmiöiden ja suhteiden olemassaolosta. Quinen (1948) mukaan käyttäessään tiettyä käsitejärjestelmää tieteellinen teoria välttämättä sitoutuu jonkinlaisiin oletuksiin maailmasta. Teoria sanoo tavallaan, että ”jos tämä teoria on tosi, niin nämä ja nämä oliot ovat välttämättä olemassa teorian mallintamassa maailmassa”.

Onttista listaa voidaan parhaiten kuvata ilmitodellisuuden ja erilaisten abstraktien käsitteiden listaksi, joka käsittää kaiken, mistä teoria puhuu. Esimerkiksi generatiivisen ja kognitiivisen kielitieteen onttinen lista on monessa kohdin toinen toisensa poissulkeva ja siten kilpaileva:

> generatiivinen kielitiede: ∃x (x on synnynnäinen kompetenssi, komputationaaliset mallit, kielen omaksuminen, universaalikielioppi jne.)
>
> kognitiivinen kielitiede: ∃x (x on opittu kielikyky, konnektionistiset mallit, kielen oppiminen, kieliopin käyttöpohjaisuus jne.).

Kilpailevatkin kieliteoriat jakavat yleensä kuitenkin kielentutkimuksen peruskäsitteistön ja olettavat, että voidaan hyväksyä olevaksi joukko perustavanlaatuisia kieli-ilmiöitä:

> ∃x (x on lekseemi, leksikalisoituminen, lause, kysymyssanat jne.).

Onttinen lista kattaa siis sen käsitteistön, jolla kuvataan jotakin olemassa olevaksi. Onttinen lista ei ole sama asia kuin vastaus metafyysisiin kysymyksiin ”Mistä oliot on tehty? Mikä on olioiden perimmäinen substanssi? Onko kaikki ainetta? Onko aine eri kuin tietoisuus? Nouseeko tietoisuus aineesta?”.

Termi *ontologia* on käytössä usein monimerkityksellinen, ja tämä aiheuttaa toisinaan arkitutkimuksessakin ristiriitaisuuksia ja hämmennyksiä teoreettiseen päättelyyn. Ontologialla voidaan tarkoittaa seuraavia seikkoja:

1) eksistenssioletuksia (eli onttisuutta),
2) olemassa olevan olion perimmäistä luonnetta (eli metafysiikkaa),
3) tutkimusalaa tai
4) kysymyksiä näistä kaikista kolmesta.

Kun kielentutkimuksessa puhutaan kielen ontologiasta, on siis mielekästä pohtia, mitä kirjoittaja viime kädessä tarkoittaa edellä kuvatuista neljästä mahdollisesta ontologian merkityksestä. Tässä johdannossa käytetään termiä *onttinen* eksistenssioletuksista (∃x), termiä *onttinen lista* joukosta todellisuuden olioita, jotka ovat olemassa jossakin teoriassa, ja termiä *metafyysinen* niistä oletuksista, jotka koskevat todellisuuden perimmäistä luonnetta. *Ontologiaksi* vuorostaan kutsutaan tässä sitä alaa, jossa tutkitaan kaikkia näitä kysymyksiä.

Yleensä jokainen kielentutkija pohtii jossakin vaiheessa kysymystä, onko havaitun puheensorinan tai luetun tekstin takana jokin kielen perimmäinen luonne. Onko kieli viime kädessä sosiaalinen normisto, jokin abstrakti järjestelmä, biologisesta rakenteestamme kumpuavaa symbolien manipulaatiota tai jotain muuta?

Ilmitodellisuuden ja perimmäisen substanssin eroa pohti jo Immanuel Kant *Puhtaan järjen kritiikissään* (1981 [1781]). Kantin mukaan *noumenon* on todellisuuden perimmäinen olemus, jonka kanssa olemme tekemisissä *fenomeenin* eli ilmiöiden välityksellä. Ainoastaan ilmiöt ovat meidän havaintomme ja kokemuksemme ulottuvilla. Kielitieteellisessä arkitutkimuksessa yleensä käsitellään kieli-ilmiöitä sitoutumatta käsityksiin kielen metafyysisestä perimmäisyydestä (ks. myös Johdannon luku 3 tk.). Tällaisessa arkitutkimuksessa siis usein riittää oman taustateorian onttisen listan hahmottaminen ja sen kysymyksen esittäminen, onko tämä lista yhteensopiva käytetyn metodin eksistenssioletusten kanssa.

Koska kielitieteen metodit on osittain sidottu kielen metafyysisiin perusteisiin, kysymys maailman perimmäisestä rakenteesta on tullut osaksi myös kielentutkimuksen metodologiaa. Esimerkiksi generatiivinen autonominen kielitiede perustaa metodinsa dualistiseen (materia vs. mieli) naturalismiin: kieli on tämän näkemyksen mukaan viime kädessä biologiaa, sillä kielijärjestelmää ohjaa mentaalinen

universaalikielioppi-moduuli (UG), joka sijoittuu aivoihin. Tällaista käsitystä vastustavat koulukunnat ovat usein sitoneet metodinsa vastakkaiseen argumenttiin. Kysymys, kuinka monesta erilaisesta substanssista maailman oliot on tehty, ei siksi ole kielitieteilijälle triviaali vaan johtaa erilaisiin tutkimuskäsitteistöihin, teorioihin ja päätelmiin.

Metafyysisten sitoumusten määrän mukaan teoriat voidaan jakaa monismiin, dualismiin ja pluralismiin. Monistit esittävät, että oliot on tehty yhdestä aineesta, esimerkiksi ideoista, fysikaalisista elementaaripartikkeleista tai jostakin neutraalista substanssista. Dualistit useimmiten esittävät, että tietoisuus ja fysikaalinen ovat kaksi erillistä substanssia (mieli vs. keho). Pluralistit taas olettavat usean keskenään erilaisen substanssin muodostavan maailman.

Sekä dualistit että pluralistit olettavat, että ajallis-paikallinen havaittu kieli (puhe, viittoma, kirjoitettu teksti) ja sen taustalla oleva abstrakti järjestelmä (normit, säännöt, UG) ovat eri asioita. Metafyysisen dualistin ja pluralistin ongelmana on useamman kuin yhden substanssin syntymekanismin selittäminen uskottavalla tavalla. Miten materiasta syntyy jotakin ei-materiaalista, kuten tietoisuus, tai toisin päin? Missä ja miten on olemassa ei-ajallis-paikallinen substanssi, jos sillä ei ole fysikaalisia ulottuvuuksia? Metafysiikassa uusien substanssien selittäminen eroaa fysikaalisesta tai biologisesta selityksestä, sillä ei-materiaalista eivät voi koskea materiaaliset selitykset. Pluralistia koskee siis raskas todistamisen taakka tieteessä, joka vaatii kertomaan, mitä kaikki uudet substanssit ovat, miten ne ovat syntyneet ja mikä niiden suhde on toisiinsa (ks. Johdannon luku 3 tk.). Jos esimerkiksi kielentutkija ryhtyy dualistiksi ja päättää, että äännökset ovat fysikaalisia ja kielen normit vuorostaan ei-fysikaalisia ja kummatkin ovat siten eri substanssia, tulee hänen vastata ainakin seuraaviin kysymyksiin: Miten normi on olemassa? Miten ei-fysikaalinen ja fysikaalinen vaikuttavat toisiinsa? Onko fysikaalisella maailmalla ja sillä (idea)maailmalla, jossa normi sijaitsee, sama tieto-oppi?

Dualistit ja pluralistit käyttävät usein termiä *emergenssi*, kun he puhuvat uusien ilmiöiden kehittymisestä. Esimerkiksi **emergenttisessä materialismissa** (ks. Niiniluoto 1990; Määttä 2000) mentaalisten ilmiöiden ajatellaan kehkeytyvän fysikaalisten lakien materiaaliselta alustalta. Mentaalinen ilmiö on tämän käsityksen mukaan osiinsa palautumaton.

Dualistin ja pluralistin on myös ratkaistava, johtavatko uudet substanssit episteemiseen pluralismiin, eli koskeeko jokaista erilaista substanssia sama vai eri tietoteoria.

### Emergenssi

Termi *emergenssi* tulee latinan sanasta *emergere* 'nousta, tulla esiin, syntyä'. Metafysiikassa termiä käytti G. H. Lewes teoksessaan *Problems of Life and Mind* (1875) tehdessään jaon uusiin emergenttisiin ilmiöihin ja niitä edeltäviin ilmiöihin. Kuten muutkin tieteelliset termit *emergenssi*-termi saa tarkan käyttömerkityksensä vasta taustateoriassaan: useita erilaisia toisistaan eriäviä *emergenssi*-käsitteitä on nykyään käytössä eri teorioissa. Käsitykset voidaan jakaa heikkoon ja vahvaan emergenssiin, joiden välinen ero selitetään usein Pierre-Simon Laplacen (1951 [1814]) kaikkitietävän demonin avulla.

Laplacen demoni tuntee kaikki universumin lait, tosiseikat, tulevat ja menneet tapahtumat atomin tarkkuudella. **Heikko emergenssi** viittaa johonkin sellaiseen kompleksisen systeemin uuteen ominaisuuteen, joka on sen systeemisen toiminnan tulos ja jonka kaikkitietävä demoni voi selittää ilmiön systeemisen toiminnan osilla ja summilla. Heikko emergenssi viittaa siis uuteen systeemin aiheuttamaan yllättävään ilmiöön, joka periaatteessa kuitenkin voitaisiin selittää tieteessä, jos tutkijoilla olisi demonin tiedot maailmankaikkeudesta.

**Vahva emergenssi** vuorostaan on uusi ominaisuus, jota koskevat totuudet eivät seuraa sen osatekijöistä tai historiaa koskevista totuuksista. Vaikka kaikkitietävä demoni tuntisi kaikki ilmiön aiheuttaneet alkutekijät, luonnon ja ihmisten aiheuttamien kompleksisten systeemien toiminnot, eivät uutta ominaisuutta koskevat todet lauseet olisi päätelmiä tästä tiedosta. Ilmiö on siis aidosti uusi jollakin sellaisella tavalla, joka on ontologisesti palautumaton alkutekijöihinsä tai osiinsa.

**Aiheesta lisää:**

Chalmers, David. 2006. Strong and weak emergence. Julkaisussa: Clayton, P. & Davies, P. (toim.) *The Re-emergence of Emergence*. Oxford: Oxford University Press, 244–256.

O'Connor, Timothy & Hong, Yu Wong. 2005. The metaphysics of emergence. *Noûs* 39:4, 658–678.

Emergenttiselle näkemykselle vastakkainen käsitys **eliminatiivinen materialismi** on saanut paljon vaikutteita myöhemmästä Wittgensteinista. Esimerkiksi Dennett (1999) esittää, että mystiset subjektiiviset kokemukset eli kvaliat eivät ole todellisia. Kvaliat, kuten kipu tai punainen, ovat merkityksellisiä subjektin tietoisuuden tai kokemuksen näkökulmasta, mutta ne saavat tämän merkityksen julkisessa jaetussa kielipelissä. Kvalioissa ja ensimmäisen persoonan perspektiivissä, jota myös kutsutaan tietoisuudeksi, ei siis Dennettin mielestä ole mitään sellaista ymmärryksen ylittävää uusien ilmiöiden emergoitumista, jota ei voisi tieteessä selittää. Tämä tarkoittaa, että Dennettin (mt.) mukaan kielen normissakaan ei ole mitään sellaista, mitä ei voisi selittää ainoastaan materiasta koostuvan maailmankäsityksen kehyksessä: linnun viserryksessä on säveliä, ja ihmisen puheessa on tavuja.

Sellaiset kielentutkimuksessa käytetyt oleelliset termit kuin *normi*, *laki*, *universaali* tai *merkitys* viittaavat johonkin ei-havaittavaan (laista ks. myös Johdannon luku 3 tk.). Ei-havaittava ei tarkoita kuitenkaan loogisesti, että jotakin uutta substanssia on välttämättä ”emergoitunut” reaalitodellisuudessa. Standardikäsityksen mukaan tieteellisen teorian käsitejärjestelmän ”yläpäässä” on abstrakteja käsitteitä, joilla ei ole havainnossa suoria vastineita. Esimerkiksi *substantiivi* tai *merkitys* ovat sellaisia abstrakteja käsitteitä, joita voi selittää vain toisten käsitteiden avulla, ei yksinomaan osoittamalla jotakin reaalista oliota. Teorian käsitejärjestelmän ”alapäässä” on vuorostaan havaintotermejä, joiden avulla teoria tarrautuu kiinni todellisuuteen. Havaintotermejä ovat esimerkiksi *lehmä*, *glottaaliklusiili* tai *pistävä*. Niiden avulla kerrotaan,

miten teoriassa olevista ilmiöistä voidaan tehdä havaintoja. Havaintotermit eivät kuitenkaan ole teoriavapaita, vaan nekin saavat määritelmänsä taustakäsitejärjestelmässä ja ovat siten teoriapitoisia. Esimerkiksi mikä tahansa kurkun kakistelu ei ole glottaaliklusiili, jos se ei ole osa tietynlaista kielellistä ilmausta ja osa kyseisen kielen äännejärjestelmää. Havaintotermi ei myöskään tarkoita, että havainnon pitäisi olla ilmiselvä nimenomaan ihmisen aistijärjestelmälle. Niin neurolingvistiikka kuin hiukkasfysiikkakin tekevät havainto-oletuksia, mutta mikrotasoa kuvaavia ilmiöitä voidaan havaita ainoastaan tutkimusteknologian eli erilaisten laitteiden avulla.

Kielentutkimuksen teoria usein nivoo *kieli*-käsitteen johonkin perimmäiseen ilmiöön eli metafysiikkaan. Chomskylle kieli perimmäisenä ilmiönä palautuu biologiaan (genetiikkaan). Siksi siinä on jotakin universaalia, jonka voi tavoittaa formaalisten järjestelmien avulla. Lakoff ja Johnson (1999) vuorostaan esittävät ajatuksen, että kieli on pääasiassa ”ruumiillista kognitiota” (engl. *embodied cognition*), joka lajivalinnan tuloksena on kehittynyt jäsentämään maailmaa kolmiulotteisen realistisesti ja aikapaikkaisen jäsentyneesti. Kielen tutkimus on siksi heidän mukaansa kognitiivisten ja ruumiillisten käsitteiden tutkimusta. Funktionalistiset teoriat vuorostaan korostavat kieltä sosiaalisena aktiviteettina tai toimintaa tuottavana prosessina. Croft (2000) kuvaa kieltä populaatioksi ilmauksia, joita tulee ajatella samalla tavoin kuin ihmispopulaation ilmiyksiköitä: ilmausten alla on *lingueme*-rakenne, ”kielen DNA”. Kielissä on siksi Croftin (mt.) mukaan sekä universaaleja piirteitä että suurta variaatiota hyvin samantyyppisesti kuin biologisissa järjestelmissä.

Kielentutkimuksessa metodin taustateoriat vastaavat siis usein sekä onttiseen että metafyysiseen kysymykseen (ks. Johdannon luku 3 tk.). Teoriat kielestä ilmaisevat eksistenssioletuksia, jolloin ne samalla ilmaisevat käsityksen kielen ja maailman luonteesta tavalla, jolla on ontologisia implikaatioita. Ilman tällaisia tutkimuskohteen olemassaoloa koskevia oletuksia tieteellisen metodin käyttöä ei voi perustella.

Kielentutkimuksessa tutkimuskohteen määritelmään liittyvä sitoumus sitoo tutkimuksen kauttaaltaan samanaikaisesti semanttiseen, ontologiseen ja metodologiseen teoriaan. Esimerkiksi sellainen käsitys,

että maailmassa on olemassa universaalia sinisyyttä tai että kieliopissa on ihmisestä riippumattomia ideoita eli universaaleja, on **käsiterealismia**. Päinvastainen käsitys puolestaan on **nominalismia**. Nominalistit siis kieltävät käsiterealismin eli sen ajatuksen, että yleiskäsitteet ovat olemassa perustavalla tavalla ihmisestä riippumattomina erillisinä olioina. Nominalistin mukaan universaaleilla ei ole vastaavuutta kielen ulkopuolisessa maailmassa. Väite, että jokin on sininen, kuvastaa nominalistin mielestä ainoastaan kielen käyttöä. *Sininen* on pelkkä nimi, jonka kieliyhteisö on antanut yhteiselle tavalle nähdä esimerkiksi meren tai taivaan heijastamat valon aallonpituudet. Nominalismi johtaa usein metodologiseen instrumentalismiin: tieteen teoreettiset termit ovat pelkkiä instrumentteja, joiden avulla tehdään ennustuksia, havaitaan ja systematisoidaan.

Tutkijan käsitys tutkimuskohdetta luokittavista käsitteistä sitoo hänet siis johonkin ontologiseen kokonaisnäkemykseen. Jos esimerkiksi kielentutkija ajattelee, että termi *genetiivi* on tutkijan työkalu mutta että todellisuudessa ei sinänsä ole olemassa erillistä oliota tai kategoriaa "genetiivi", hän saattaa olla nominalisti ja metodologinen instrumentalisti. Jos kielentutkija vuorostaan tunnustautuu käsiterealistiksi, hän uskoo, että termi *genetiivi* kuvaa jotakin luonnollista kielellistä luokkaa, joka on olemassa reaalimaailmassa silloinkin, kun kukaan ei ole sitä luokittamassa. **Konseptualismi** on realismin ja nominalismin välissä oleva kanta, joka olettaa, että yläkäsitteet ovat ihmisten mielessä. Konseptualistin mukaan havainto maailmasta on käsitteellistämistä. Maailma ei konseptualismin mukaan ole erillinen käsitejärjestelmästämme vaan osa sitä. Jotta mitään voisi sanoa miksikään, esimerkiksi *genetiiviksi*, ihmisen on ensin käsitteellistettävä havaintonsa. Käsitteitä on siten ainoastaan ihmismielessä, ja siksi konseptualistin mukaan on mieletöntä kysyä mitään maailman käsitteellisestä rakenteesta. Mielekkäitä ovat vain ne kysymykset, jotka kysyvät, miten käsitteellistämme maailmaa.

**Konventionalistit** taas ajattelevat, että kielet, tieteelliset teoriat ja yläkäsitteet ovat sopimuksia. Monet nykysuuntaukset, jotka edustavat ontologista pluralismia tai käsitystä, että maailman kuvaamisen kieliä on useita, perustavat käsityksensä Carnapin (1934) konventionalismiin. Carnapin (mt.) mukaan todellisuuden luonnetta koskevat kysymykset

riippuvat kielen rakenteesta. Kielen merkitykset ovat sopimuksenvaraisia. Carnapia seuraavan saksalaisen nykyfilosofin Markus Gabrielin (2015) mukaan voimme eri kielissä kuvata yhden reaalimaailman sijaan useita erilaisia merkityskenttiä. Merkityskenttä tarkoittaa sitä, että kvantteja on olemassa David Deutschin kvanttifysiikan teoriassa, noitia on olemassa Grimmin saduissa, skeemoja on olemassa kognitiivisessa kieliopissa jne. Kun muut metafysiikan teoriat määrittävät kysymyksen eksistenssistä kielen ja reaalimaailman väliseksi kysymykseksi, konventionalisti sijoittaa sen tekstin tai kontekstin sisäiseksi kysymykseksi.

Suomessa varsin paljon huomiota saaneen, Niiniluodon (1999) kehittämän **kriittisen realismin** näkemys on vuorostaan se, että todellisuudessa on käsitteellistämisestä ja kielestä riippumattomia piirteitä, joista voimme saada tietoa vain epäsuorin keinoin selvittämällä asioita parhaamme mukaan tieteen tai muun tietämisen tavan kautta. Yläkäsitteet viittaavat todellisuuteen. Niiniluoto puolustaa käsitystä, että tieteellisen teorian teoreettiset käsitteet, kuten *atomi* tai *genetiivi*, kuvaavat jotakin reaalitodellisuudessa. Se että teorioita voi testata todellisuuden ja aineistojen valossa, osoittaa tieteellisen realistin mukaan sen, että teoreettiset käsitteet eivät ole pelkästään instrumentaalisia. Myös Lakoff ja Johnson (1999) pitävät omaa merkitysteoriaansa realistisena, eli todellisuudessa todella on prototyyppejä ja metaforia, niin kielessä kuin ihmisen mielessäkin.

1900-luvulla tieteenfilosofiassa pohdittiin laajasti sitä, onko tieteellisen teorian abstrakteilla eli ei-havaittavilla käsitteillä vastinetta eli referenttiä todellisuudessa. 1900-lukua voidaankin kutsua *teoreetikon dilemman* (Hempel 1958) vuosisadaksi.[7] Debattia seurannut nykykäsitys on, että kaikki termit ovat teoriapitoisia eli termit sisältävät merkitysrakenteita taustateoriastaan (Tuomela 1973). Tämä tarkoittaa sitä, että havaintotermin ja abstraktin termin välillä ei ole dikotomista eroa. Niiden rakenne vain on erilainen.

Edellä huomattiin, että käsitys siitä, mihin teoreettiset termit viittaavat, implikoi totuusteoriaa. Totuuden **korrespondenssiteoria** eli käsitys, että teoria kuvaa todellisuutta, on sovellettavissa ainoastaan realistisessa kehyksessä, joka hahmottaa kielestä riippumattoman maailman. Instrumentalistin soveltamat totuuden **koherenssiteoriat** vuorostaan esittävät,

että koko käsite *todellisuus* on hylättävä, sillä meillä ei ole käsitteistä riippumatonta suhdetta todellisuuteen. Koherenssiteorioiden taustalla on käsitys kielen semantiikasta: kieli voi viitata vain muihin kielellisiin merkkeihin, ja merkitys muodostuu vain suhteessa kielelliseen kontekstiin. Koherenssiteorioiden mukaan voidaan periaatteessa puhua vain erilaisten todellisuutta kuvaavien teorioiden välillä esiintyvästä koherenssista, ei ihmisestä riippumattomasta reaalitodellisuudesta itsestään. Mikäli jokin tieteellinen väite on ristiriidassa hyväksyttyjen muiden tieteellisten väitteiden kanssa, sitä ei hyväksytä tieteelliseksi tiedoksi. Mikäli väitteiden ja teorioiden välillä on suuri koherenssi, voidaan puhua tieteellisen tiedon karttumisesta. **Konsensusteorioiden** mukaan ei voida puhua edes koherenssista, sillä viime kädessä tiedeyhteisö on se, joka päättää siitä, mikä on mielekäs tieteellinen lause ja mikä ei. Näin ollen tulisi puhua vain siitä, mitä tiedeyhteisö subjektina päättää uskoa todeksi. Konsensusteorioita motivoi näkemys, että kielelliset merkitysjärjestelmät ovat mielivaltaisia, kulttuurikohtaisia ja satunnaisia. Instrumentalismin ja realismin sekä koherenssi- ja korrespondenssiteorioiden välistä jännitettä ilmaistaan myös termien *konstruktivismi* ja *fundamentalismi* avulla. Fundamentalismi on realistisiin käsityksiin liittyvä ajatus siitä, että tieto on perustavanlaatuista ja liittyy maailmaan, kun taas konstruktivismin teoriat yleensä esittävät, että tieto on ihmisen rakentamaa, konstruoitua.

Totuusteorioiden taustalla on erilainen käsitys semantiikasta:

- **Realistiset teoriat** olettavat, että kieli viittaa johonkin todelliseen maailmassa. Siksi käytössä on usein jonkinlainen totuuden **korrespondenssiteoria**.
- **Instrumentalistiset teoriat** vuorostaan olettavat, että kieli on sopimuksenvaraista. Käytössä on siksi jokin totuuden **koherenssi**- tai **konsensusteoria**.

Viime vuosisadalla keskustelua käytiin kielen ja sen referenttien eli sen ilmitodellisuudesta poimimien tarkoitteiden suhteesta. Nykydebatin kohteena on havaintokielen ja tieteen kielen välinen ero. Tämä dilemma

voidaan kiteyttää kysymykseen siitä, onko havaittuja ilmiöitä lainkaan olemassa. 2000-luvun metafyysistä keskustelua voidaan siksi kutsua *havaitsijan dilemmaksi*. Thomassonin (2007) mukaan nykytutkimuksessa on enemmän sellaisia koulukuntia, jotka esittävät, että mitään ei ole olemassa havaitsijasta riippumatta. Nämä koulukunnat kieltävät jopa tavallisten arkipäivän olioiden olemassaolon: pöytiä tai Matteja ei ole olemassa. Tähän **nihilistiseen** tai **eliminativistiseen** suuntaukseen on syynä tieteellisen maailmankuvan muutos. Ongelmaksi hahmottuu se, että tieteellisen maailmankuvan ja havaintomaailman välissä näyttää olevan ylittämätön kuilu. Tieteen kuvaama fysikaalinen todellisuus ja ihmisen havaitsema todellisuus tuntuvat kuvaavan hyvinkin erilaisia asioita: kuvauksissa ei ole käytössä mitään yhteisiä termejä.

Quinen (1948) tavoite tunnetussa kirjoituksessaan ”On what there is?” oli onttisen kysymyksen muotoa jäsentämällä pyrkiä siihen, että ontologia toimisi rakentavana osana tieteellistä teoriaa ja tieteellistä tutkimusta. Paradoksaalisesti juuri tämä näkemys ontologiasta osana tieteellistä teoriaa on aiheuttanut tutkimussuunnan, jossa on yllättävän vaikea perustella arkihavaintomme olioiden olemassaoloa (ks. Thomasson 2007). Jos hyväksymme tieteellisen maailmankuvan esittämän todellisuuden ja sen, että se on jollakin tapaa todempi kuin aistihavaintomme, silloin tämä johtaa kysymykseen, mitä aistienvarainen arkihavainto sitten ilmaisee. Tämän kysymyksen voi sopeuttaa kielentutkimukseen seuraavassa muodossa: jos kielentutkimus tavoittaa todellisemman kuvan kielestä kuin arkihavainto, mitä arkihavainto kielestä sitten ilmentää? Se että esimerkiksi kuulee selvästi toisen ihmisen puhuvan ja jopa havaitsee puheessa erilaisia sijamuotoja, ei tämänhetkisen metafysiikan alan tutkimuksen mukaan ollenkaan todista, että kieltä on olemassa.

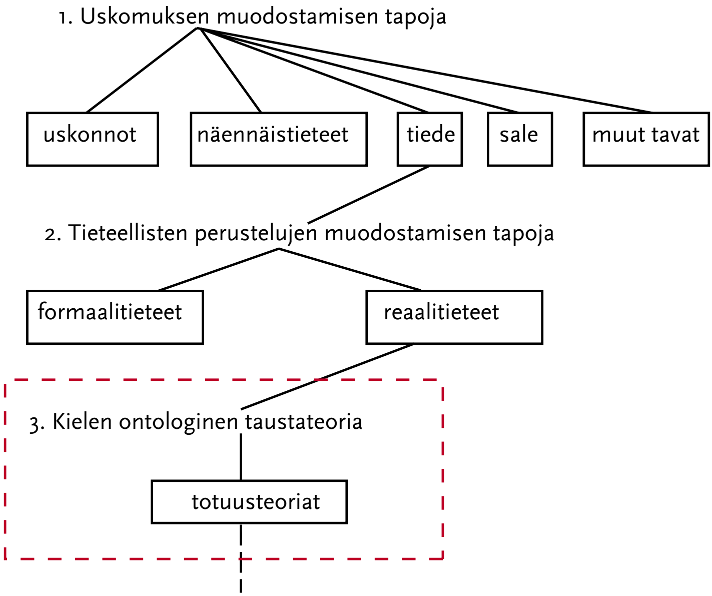
Metafysiikan nihilistinen suuntaus on johtanut viime vuosikymmeninä lisääntyvään määrään metametafysiikan koulukuntia, joissa on alettu tutkia ontologian ja metafysiikan alan tutkimuskäsitteistöä (ks. Chalmers, Manley & Wasserman 2009). Realismia eli sitä käsitystä, että on olemassa jotakin mielestä riippumatta, puolustetaan siksi 2000-luvulla hyvin usein antirealististen metaontologisten ja nihilististen metametafyysisten käsitysten kautta. Nämä metametapositiot kieltävät tai kyseenalaistavat osan ontologisesta tutkimuksesta tai

kaiken sitä koskevan tutkimuksen. Ontologisen ja metaontologisen tutkimuksen tasojen välistä eroa voi kuvata seuraavasti: Ontologi kysyy, onko maailmassa kieltä, johon idealisti vastaa ”kyllä” ja nominalisti ”ei”. Metaontologiassa taas kysytään, onko olemassa sellaista objektiivista näkökulmaa kieleen, jonka avulla saisimme selville, onko maailmassa olemassa kieltä. Metaontologinen realisti vastaa tähän kysymykseen ”kyllä” ja metaontologinen antirealisti ”ei” (ks. Chalmers 1996). Viime vuosisadalla kielentutkija on voinut vedota sellaisiin filosofisiin auktoriteetteihin kuin Wittgensteiniin, Quineen tai Chomskyyn valitessaan metodin sillä perusteella, että uskoo kielen olevan jotakin erityistä substanssia tai vastustaa jotakin toista koulukuntaa, koska pitää sen käsitystä kielen ”todellisesta luonteesta” virheellisenä. 2000-luvulla kielentutkijan filosofinen konteksti on hyvin toisenlainen. Kielifilosofiankin perusta on hieman horjunut, koska metafysiikan tutkimus on hajonnut kahteen kenttään.

Olipa tutkijan ontologinen sitoumus mikä tahansa, tutkijan tieteellisen teorian käsitejärjestelmän tulee olla joka tapauksessa sisäisesti koherentti. Koherentissa ja loogisessa tieteellisessä teoriassa väitelauseiden tulee säilyttää onttiset, metafyysiset ja kategorioita koskevat oletukset. Tämä tarkoittaa seuraavaa:

1. Vastaus onttiseen kysymykseen (Mitä olioita maailmassa on?) on teoriassa koherentti ja looginen.
2. Vastaus metafyysiseen kysymykseen (Mistä maailma on tehty?) on koherentti ja pysyy tutkimuksen kuluessa vakiona. Suhtautumista tai asemoitumista eri näkökantojen, esimerkiksi monismin tai pluralismin välillä, ei vaihdella.
3. Käsitejärjestelmä on sisäisesti koherentti: ei esimerkiksi käytetä emergenttistä terminologiaa, jos on sitouduttu eliminatiiviseen materialismiin.
4. Kategoriaoletukset pysyvät samoina teoriassa. Tämä tarkoittaa, että käsitteiden intensio ja ekstensio pysyvät teoriassa muuttumattomina. Tämä tarkoittaa myös sitä, että alakäsitteiden määritelmien ja yläkäsitteiden välillä on merkityskoherenssi: yläkäsitteen intensio on osa alakäsitteen määritelmää.

Edellä on tutkittu sitä, miten kielitieteessä tutkimuskohteen eli kielen määritelmä vaikuttaa niin metodologiaan kuin tieto-oppiin ja totuusteoriaan. Huomattiin, että käsitys semantiikasta implikoi oletuksia kielen ja maailman suhteesta ja siten viime kädessä määrittää totuusteoreettiset oletukset. Tämä suhde näkyy kuviossa 10.




Kuvio 10. Kielen oletettu ontologia implikoi totuusteoriaa.

## 2.5. Metodologiset ehdot

Edellisessä luvussa on tarkasteltu sitä, millä tavoin tieteellinen tutkimus perustuu käsityksiin tutkimuskohteesta ja siihen, miten kohteesta saadaan tietoa. Tässä luvussa tarkastellaan sitä, millaisia erilaisia vastauksia metodologisiin kysymyksiin on annettu ja millä tavoin annetut vastaukset ovat ajan myötä kehittyneet. Tieteellistä prosessia pohtineet tieteenfilosofit ovat perustaneet käsityksensä oikeasta metodologisesta ehdosta tai periaatteesta esimerkiksi näkemyksiin siitä, että tieteen päämääränä

on yhtenäinen maailmankuva tai että jokin tietty selitysmalli on ainoa oikea tieteellisen selityksen rakenne. Oheiseen tietolaatikkoon on kiteytetty niitä kysymyksiä, jotka ovat tieteenfilosofien päättelyssä johtaneet asettamaan tietyn metodologisen ehdon tieteelliselle tutkimukselle.

**Metodologisia kysymyksiä**

- Onko tieteellä yksi yhteinen päämäärä?
- Onko tieteellä jokin yksi (tai useampi) yhteinen piirre?
- Liittyvätkö tieteelliset tiedot toisiinsa muodostaen yhden yhtenäisen kuvan todellisuudesta?
- Onko tiede objektiivista vai arvolatautunutta?
- Millaisia selityksiä ja kuvauksia tieteessä hyväksytään?
- Ovatko termit, teoriat tai tutkimusalat palautettavissa toisiinsa?

Metodologinen optimisti Francis Bacon (1561–1626) uskoi (1620), että hän oli kehittänyt vedenpitävän induktiivisen tekniikan, jota voisi soveltaa mekaanisesti ilmiöön kuin ilmiöön, jolloin tuloksena olisi tieteellisiä totuuksia. David Hume (1711–1776) huomasi (2016 [1748]) kuitenkin, että Baconin ehdottaman induktivismin ongelma on se, että emme koskaan voi tyhjentävästi havainnoida kaikkia yksittäistapauksia: jossain on voinut olla joskus musta joutsen, joka kieltää lauseen ”(kaikki) joutsenet ovat valkoisia”.[8] Toisekseen lakia ei voi johtaa yksittäisistä ilmiöistä. Baconin induktivistinen teoria ei siis ollut sellainen tieteellinen prosessointilaite, jonka yhteen päähän saattoi työntää havaintoja ja toisesta päästä tuli ulos luonnonlakeja.

**Naiivi ja kehittynyt induktivismi**

**Naiivin induktivismin** mukaan tutkimus etenee havainnoista yleistyksiin, joita sitten testataan uusien havaintojen avulla. Naiivi induktivismi on ankaraa empirismiä, sillä se esittää, että havainto on riittävä tiedon lähde ja tarjoaa varman perustan.

**Kehittyneen induktivismin** mukaan tutkija määrittelee tutkimusongelman ja aineiston tarpeeksi suppeaksi siten, että kysymykseen saadaan vastaus yhdessä tutkimuksessa. Tutkimuksessa ei pyritä yleisiin lakeihin vaan tiettyä populaatiota tai aineistoa koskevaan rajoitettuun yleistykseen.

**Aiheesta lisää:**

Hacking, Ian. 2001. *An Introduction to Probability and Inductive Logic.* Cambridge: Cambridge University Press.

Auguste Comten (1798–1857) mukaan Baconin ja Humen analyyttiset ajatukset tieteestä järjestelmällisenä metodina heijastivat ihmiskunnan ajattelussa tapahtunutta positiivista edistystä. Tätä edistystä hän kuvasi kolmen vaiheen kautta:

1) Ensimmäisenä oli teologinen vaihe, jossa luonnonilmiöt ymmärrettiin yliluonnollisten olioiden aiheuttamiksi.
2) Metafyysisessä vaiheessa ihmiskunta alkoi käyttää abstrakteja käsitteitä.
3) Positiivisessa vaiheessa ihmiskunta ymmärsi, että tietoa voi saada vain havaittavista tosiasioista.

**Positivismi** tarkoitti Comtelle positiivista asennetta tieteeseen yhteiskunnassa, ja hän toivoi, että esimerkiksi päättäjät perustaisivat päätöksensä uskonnon sijaan tieteelliseen tietoon. Valistusajan positivismi oli maltillista empirismiä ja tieteellistä optimismia, ja se eroaa sitä seuranneesta ankaran reduktiivisesta 1900-luvun alun loogisesta positivismista eli loogisesta empirismistä. Kumpaakin tieteenfilosofista koulukuntaa kuitenkin yhdistää ajatus, että tieteellä on yksi yhtenäinen kriteeri ja tieteellinen tieto yhdistyy yhtenäiseksi tieteelliseksi maailmakuvaksi kaikesta olevasta. Kumpikin positivismi oli siis monistinen näkökanta. **Metodologiselle monismille** on tyypillistä olettaa,

että tieteellä on jokin yksi yhdistävä piirre tai ehto ja että viime kädessä kaikki tieteellinen tieto antautuu yhdeksi yhtenäiseksi maailmankuvaksi ihmisen ymmärrykselle.

Looginen empirismi, joka tunnetaan myös Wienin piirinä, oli hyvin lyhytkestoinen koulukunta 1920-luvulla ennen toista maailmansotaa. Sen vaikutus tieteenfilosofiaan on kuitenkin ollut huomattava, sillä siihen kuului suuri osa tuon ajan merkittävimmistä filosofeista: Rudolf Carnap, Kurt Gödel, Otto Neurath, Moritz Schlick, Philipp Frank, Karl Menger ja Hans Hahn. Useat filosofit, kuten suomalainen Eino Kaila, assosioituivat piiriin välillisesti. Sodan alkaessa usea piirin jäsen pakeni Yhdysvaltoihin, mikä johti siihen, että piirin ajatuksia alettiin ilmaista englanniksi ja ne levisivät laajemman yleisön tietoisuuteen. Piirin rooli tieteenfilosofiassa ja maininta tässäkin yhteydessä ei perustu siihen, että se olisi esittänyt oikeaksi oletetun käsityksen tieteestä – päinvastoin. Kun viimeiseltä loogiseksi empiristiksi tunnustautuneelta Alfred Ayerilta kysyttiin 1970-luvulla, mikä kovasti haukutussa loogisessa empirismissä oli niin väärää, hän vastasi, että suuntauksen kaikki premissit olivat virheellisiä. Koulukunnan ääriempiristiset ajatukset toimivatkin eräänlaisena ponnahduslautana toisenlaiseen ajatteluun. Liikkeen esittämien ajatusten vastustaminen muodostaa keskeisellä tavalla 1900-luvun tieteenfilosofisen kirjallisuuden punaisen langan (ks. Raatikainen 2004).

Loogiset empiristit olivat esimerkiksi vaatineet, että teoreettiset termit on palautettava havaintokieleen ja että jos käsitettä ei voisi kääntää välittömästi havaittavaan, se olisi hylättävä. Tätä kutsuttiin *käännettävyysteesiksi*. Tätä ja muita koulukunnan esittämiä teesejä tutkittiin laajalti sotien välisessä ja vielä sen jälkeisessä tieteenfilosofiassa. Yksi toisensa jälkeen ne osoittautuivat mahdottomiksi. Tieteellistä teoriaa ei voi pelkistää havaintoon. Nykytutkimuksessa tämä tarkoittaa sitä, että termi *empiirinen* ei ole käytössä enää tässä reduktiivisessa merkityksessä: empiirinen ei siis tarkoita, että kaikki abstrakti olisi palautettava havaittavaan. Empiirisellä näytöllä tarkoitetaan nykyään yleensä laajemmin kaikkea sellaista ensi käden informaatiota, joka voi antaa tukea jollekin väitteelle tai kieltää sen. Siten esimerkiksi keskustelunanalyysin data, erilaiset tekstit, korpukset tai kyselyt voidaan käsittää empiiriseksi todistusaineistoksi. (Termin *empiria* käytöstä kielitieteessä ks. Konstenius 2014.)

Loogisen empirismin eräs versio, **operationalismi**, vaati, että termit käännetään mittausformuloiksi. Sekään ei lopulta kantanut kovin pitkälle: eri mittareilla saadut tulokset eivät terveen järjen mukaan kerro eri olioista. Hiekkakello, digitaalinen kello ja mekaaninen kello eivät viittaa eri aikoihin. *Operationalisoinnilla* tarkoitetaan nykyään yleensä vain sitä, että yksityiskohtaisesta teoreettisesta termistä muodostetaan pelkistetympi testattava määritelmä empiiriseen tutkimukseen. Esimerkiksi Eleanor Roschin (1973) termi *prototyyppi* voidaan määritellä kyselyä varten 'kategorian parhaaksi esimerkiksi', jolloin vastaajilta kysytään heidän käsityksiään esimerkiksi värien, huonekalujen tai pelien edustavimmista esimerkeistä. Käsite *prototyyppi* ei ole synonyyminen hyvän esimerkin kanssa, mutta tutkimusasetelmaa varten tällainen köyhä operationaalinen määritelmä on välttämätön.

Loogisen empirismin **verifikationismi** eli ajatus siitä, että tieteelliset termit yksinkertaisesti todennetaan kokemuksen, havaintotermien tai operationaalisten määritelmien avulla, hylättiin myös. Karl Popper (2002 [1934]) esitti sen vastapainoksi **falsifikationismin**, vaatimuksen siitä, että tieteellisen teorian tulee tehdä sellaisia havaintomaailmaa koskevia ennustuksia, joiden perusteella voidaan päätellä, onko teoria tosi vai ei. Ennusteiden tulisi lisäksi olla sellaisia, että ne voidaan kääntää kieltolauseiksi, sillä tämä kieltolauseiden joukko kertoo, missä olosuhteissa teoria ei ole tosi, eli mikä teorian falsifioi (Popper 2002 [1934], 95).[9] Nykyään ajatellaan, että tutkimuksessa voidaan saavuttaa hypoteesia tukevaa evidenssiä tai falsifioida osia teoriasta. Koko taustateoriaa ei yleensä voi kaataa yhden tutkimuksen tuloksin.

Tieteelliset teoriat ovat oman aikansa käsitejärjestelmän osia, ja näitä käsitejärjestelmiä parannetaan vain niiltä osin, kuin ne osoittautuvat puutteellisiksi. Niitä ei hylätä ennen kuin jokin kilpaileva teoria on osoittautunut selvästi kokonaisvaltaisesti kelvollisemmaksi. **Naiivi falsifikationismi** on nimensä mukaisesti yksinkertainen mutta myös väärä tulkinta Popperin tieteenfilosofiasta. Se esittää virheellisesti, että jokainen tieteellinen teoria on operationalisoitava kokeelliseen tutkimusasetelmaan (Kitcher 1983, 44). Jos näin todella olisi, Einsteinin suhteellisuusteoria tai muut teoreettiset tietojäsennykset eivät olisi tieteessä luvallisia (Couvalis 1997, 69–71).

Nykyään myös Popperin falsifikationismia pidetään liian ankarana, sillä se johtaa siihen, ettei tieteessä voi hyväksyä rajoittamattomia eksistenssilauseita, esimerkiksi ”on olemassa joutsenia” (ks. Niiniluoto 1980b). Vuosisadan puolivaiheessa positivismin korvasi **instrumentalismi** eli käsitys siitä, että teoreettisilla termeillä oli käyttöä teorioissa. Instrumentalismin mukaan teoreettiset termit, kuten *elektroni* (fysiikassa) tai *särkyvä* (kryptografiassa tiedon piilottamisesta), ovat hyödyllisiä, mutta ne eivät viittaa mihinkään todelliseen.

**Naiivi ja kehittynyt falsifikationismi**

**Naiivin falsifikationismin** mukaan tutkimuksessa falsifioidaan eli pyritään kumoamaan lauseita eli hypoteeseja.

Premissi: Jos hypoteesi H on tosi, niin silloin saamme havaintotuloksen h.

Falsifikaatio: Havaintotulosta h ei esiinny.

Päätelmä: Hypoteesi H ei ole tosi.

**Kehittyneessä falsifikationismissa** koetellaan teorioita, jotka voivat saada vain osittaista konfirmaatiota tai vastaevidenssiä. Tutkimusasetelma tulisi rakentaa siten, että saadaan mahdollisimman suuri mahdollisuus vasta-argumentille.

Postpositivismin kehitykseen on vaikuttanut pitkälti Sellarsin (1963) huomio, että havainto ei ole puhdasta. Nykypsykologia on tukenut tätä käsitystä, ja se on osoittanut kiistatta, että havaitseminen on teoriapitoista, Wittgensteinia (2009 [1953]) mukaillen ”näkemistä jonakin”. Ei siis ole olemassa käsitteistä vapaata puhdasta havaintoa todellisuudesta siinä mielessä kuin positivistit olettivat. Tämä tarkoittaa sitä, että ero teoreettisen ja empiirisen tutkimuksen välillä on kaventunut ja hämärtynyt.

Loogisen empirismin jälkeen alettiin keskustella erityisesti Hempelin ja Oppenheimin (1948) deduktiivis-nomologisesta mallista (DN-malli),

joka on yksi versio John Stuart Millin (1806–1873) esittämästä **peittävän lain mallista** eli **subsumptioteoriasta** (1906 [1875]). Subsumptioteoria on tieteellisen selittämisen malli, joka vastaa kysymykseen, miksi ilmiö tapahtui. Mallissa selitettävä ilmiö seuraa yleisistä laeista ja alkuehdoista.

Selitysongelma: Miksi järvi on jäätynyt?

Yleinen laki: Vesi jäätyy 0 celsiusasteessa.

Päätelmä: Järvi on vettä, ja ilman lämpötila on alle 0 celsiusastetta.

Subsumptioteorian yksi ongelma on, että tieteellinen tieto ei aina ole selittävää. Osa tieteellisestä tiedosta on kuvailevaa, analyyttista, systematisoivaa ja syntetisoivaa. Esimerkiksi usein jonkin kielen referenssikielioppi on tiedollinen systematisaatio, ei selitys siitä, miksi kieli on sellainen kuin on. Tällainen kuvaileva, tulkitseva ja systematisoiva tutkimus on kuitenkin oleellista perustutkimusta useimmilla tieteenaloilla. Subsumptioteorian tutkimus johti siis päätelmään, että tiede on monimuotoisempaa kuin aikaisemmin oli luultu. Keskenään hyvinkin erilaiset tietoesitykset ovat tiedettä.

Subsumptioteorian toinen ongelma oli myös käsitys kausaalisuudesta lainomaisena säännönmukaisuutena. Tämä kausaatio-käsitys on nykyfilosofian mukaan kestämätön, ja sen on korvannut käsitys kausaatiosta manipulaationa: A aiheuttaa B:n. Kausaation nykymääritelmään kuuluu myös usein kontrafaktuaalinen ehto: ilman B:tä ei tapahtuisi A:ta (ks. Raatikainen 2007). Kausaation määritelmällä on ollut suuri vaikutus siihen, miten sekä fysikaalisen maailman että sosiaalisen maailman suhteita kuvataan ja voidaan ymmärtää. Kausaatio-käsitteen modernin merkityksen ymmärtäminen on myös oleellista kielen teoriassa: kausaatio ei siis tarkoita luonnonlain kuvaamista vaan aiheuttamisen kuvaamista.

Nancy Cartwrightin (1983) mukaan subsumptioteorian sovelluksia, kuten sen modernimpia versioita Suppen (1974) probabilistista kausaatiota, Salmonin (1971) tilastollista relevanssia tai Hanssonin (1974) kontekstuaalista mallia, vaivaa sisäänrakennettu illuusio luonnon lainomaisuudesta. Todellisia kausaalisia luonnonlakeja on vain muutama, loput ovat niin sanottuja *ceteris paribus* -yleistyksiä. *Ceteris paribus* (lat.)

tarkoittaa muiden asioiden pysymistä muuttumattomina. Ilmausta käytetään tieteenfilosofiassa, kun halutaan viitata siihen, että jokin ulkopuolinen tekijä saattaa vaikuttaa kausaalisen prosessin lopputulokseen. Todellisuus on sotkuista, ja fysiikankin lait ovat *ceteris paribus* -idealisaatioita, jotka kertovat, miten asiat olisivat, jos mikään muu seikka ei vaikuttaisi niihin. Cartwrightin (mt.) mukaan ei ole mitään periaatteellista syytä, miksi Jumala tai luonto ei olisi voinut tehdä vain muutaman lain ja rajoituksen ja sitten väsyä projektiinsa ja jättää todellisuuden perimmäisiä suhteita sotkuisiksi. Ajatus siististä, yhdestä lainomaisesta yhtenäisestä matemaattisesta maailmasta, joka on kaikkien ilmiöiden taustalla, on liian idealistinen. Cartwrightin (mt.) mukaan – jos todella rohkenemme katsoa, mitä fysiikassa tehdään ja millaisia sen teoriat tosiasiassa ovat – fysiikan teoriat muistuttavat enemmän ihmistieteiden sotkuisia malleja, joiden käsitteet ja väitteet ovat usein keskenään ristiriidassa. Kielentutkimuksen tutkimuskäsitteistön vertaaminen johonkin ideaalina pidettyyn luonnontieteeseen, jossa löydetään puhtaita luonnonlakeja, ei siis ole nykykäsityksen mukaan kovin mielekästä.

Edellä olemme havainneet, että perimmäistä todellisuutta koskevat käsityksemme ohjaavat tieteen tekoa: Tuleeko tieteellisen tiedon yhdistyä yhdeksi yhtenäiseksi maailmankuvaksi? Onko maailma kausaalisesti sotkuinen paikka? Cartwrightin (1983) mukaan luonnontieteen selitykset eivät ole sen lainomaisempia kuin ihmistieteenkään: kaikki selitykset tieteessä ovat *ceteris paribus* -selityksiä. Cartwrightin (mt.) käsitys tieteestä on siis monistinen mutta ei perustu tieteellisen tiedon yhdistämispyrkimykseen.

Metodologinen monismi voi siten perustua keskenään hyvinkin erilaisiin argumentteihin: käsityksiin tieteen yhtenäisyydestä, tieteellisyyden kriteeriin, selityksien johonkin piirteeseen tai todellisuuden luonteeseen. Raatikainen (2007) on muiden ohella esittänyt käsityksen, että metodologista monismia on mahdollista kannattaa ja esimerkiksi vaatia, että ihmistieteiden *kontrafaktuaalisen kausaation* käsitettä sovelletaan myös luonnontieteisiin. Jo Kuhn (1971) kiinnitti huomiota siihen, että *lainomaisen kausaation* käsite oli häviämässä fysiikan teorioista (Loewer 2001). Jotkut fysiikan perusteita pohtivat nykyfilosofit ovatkin väittäneet, että *lainomaisen kausaation* käsite ei ole ollenkaan sovellettavissa fundamentaaliseen

fysiikkaan (ks. Latham 1987; Redhead 1990; Field 2003; Hitchcock 2007; Elga 2007). Fysiikan kausaation käsite olisi siten enemmän ihmistieteiden kausaatio-käsitteen kaltainen kuin, mitä viime vuosisadalla oletettiin.

Metodologiset dualistit, kuten Charles Taylor (1971), esittävät yleensä, että luonnontieteillä ja ihmistieteillä on erilaiset metodit; toinen on havainnoiva ja toinen on ymmärtävä eli **hermeneuttinen**. Ihmistieteitä erottaa se merkitys, jonka ihminen antaa ilmiöille ja joka sitten viime kädessä ohjaa hänen toimintaansa. Käsitys siitä, että ihmisen luoman maailman merkitysten tulkinta perustuu ymmärrykseen, on peräisin erityisesti Wilhelm Diltheyltä (1982 [1860]), ja ajatusta ovat kehittäneet edelleen Hans-Georg Gadamer (1976) ja Jürgen Habermas (1968) kirjoituksissaan. Hermeneutiikka on viime aikoina saanut myös uusia tulkintoja ja sovelluksia uusissa tieteellisissä yhtymäkohdissa, kuten informaatiosysteemien ja digitaalisen humanismin risteyskohdissa (Scheuermann & Kroesze 2017) sekä ihmisten luomien merkitysmaailmojen, esimerkiksi virtuaalisten tilojen, ontologisten kielien ja informaatioekologian toimintojen, ymmärtämisessä. *Hermeneutiikka* käsitteenä on myös inkorporoitunut kognition ja kommunikaation ymmärtämiseen: neuraalihermeneutiikka (Friston & Frith 2015) on osa monia malleja, jotka selittävät sosiaalista kognitiota, ymmärtämistä ja yhteistoimintaa (Gallagher & Allen 2016).

Myös dualistit voivat siis perustella näkemyksensä keskenään erilaisin argumentein. Esimerkiksi von Wright (1999 [1971]) on metodologinen dualisti, joka esittää, että ihmistieteissä ei voida puhua kausaatiosta samassa mielessä lainomaisena lakina kuin luonnontieteissä. Ihmisten kohdalla tulee käyttää teleologisia selityksiä: intentio, motivaatio tai uskomus selittää tapahtumia. Ajatus on johdettu Aristoteleen (1994) praktisesta syllogismista: A uskoo, että tekemällä B:n hän saavuttaa C:n.

Ymmärtäminen inhimillisenä aktiviteettina implikoi jonkinlaisia arvoja. Tämä tarkoittaa sitä, että hermeneuttinen traditio on ristiriidassa weberiläisen tiedekäsityksen kanssa. Max Weberin (1949) mukaan tiede ei voi kertoa meille, mihin ihmiskunnan tulisi pyrkiä, vaan ainoastaan, miten sinne pääsee. Weberiläinen tiedekäsitys erotti tieteen objektivismin ja arvot. Hermeneutiikan isä Habermas (1970) vuorostaan esitti, että nämä ovat tieteessä saman kolikon kaksi puolta: ihmistieteet ovat

arvolatautuneita, mutta niin on muukin tiede. Habermasin mukaan näin välttämättä aina onkin: tiede on inhimillinen aktiviteetti, ja sen tulisikin edistää vapautta, hyvinvointia ja rauhaa.

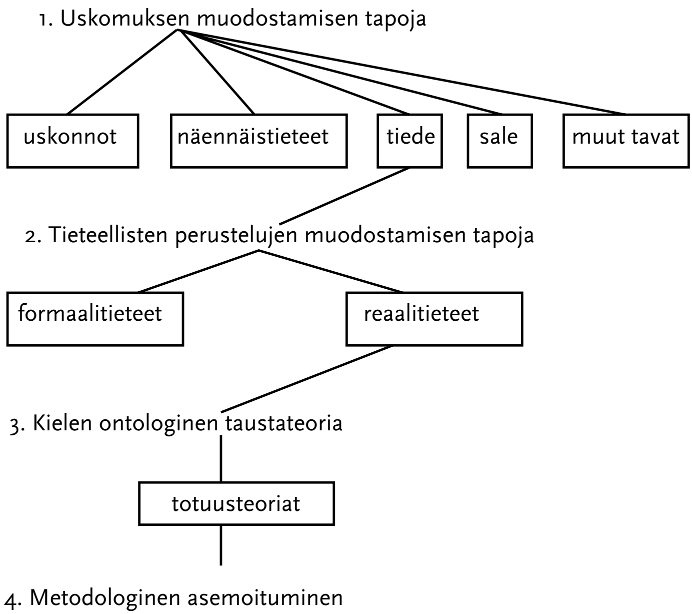
Habermasin käsitystä voidaan kutsua positiiviseksi. Pessimistisen version arvolatautuneesta tieteestä esittivät ranskalaiset poststrukturalistit. Michel Foucault, Jacques Derrida, Gilles Deleuze, Jean Baudrillard sekä Julia Kristeva kritisoivat käsitystä, että tieteellinen tieto olisi pelkästään objektiivista ja neutraalia: rationaaliseen järkeen vetoaminen ei poista velvollisuutta tuoda esiin tieteen valtapyrkimyksiä. Heidän käsityksensä yhdistää Jacques Lacanin psykoanalyysin, marxilaisen dialektiikan ja Friedrich Nietzschen nihilistisen käsityksen: tiede toimii myös dominanssin välineenä.

Poststrukturalistisen liikkeen mukana esimerkiksi Judith Butler aloitti 1970-luvulla feministisen tiedekritiikin ja herätti kysymyksen, onko tiede osa patriarkaalista valtarakennetta. Feministinen tieteentutkimus tutkii, missä määrin neutraali sukupuoli (esim. *man* tai *human*) viittaa tutkimuksissa miehiin, missä määrin naisten perspektiivi on mukana tutkimuksessa tai millä tavoin esimerkiksi metaforat ovat tieteessä maskuliinisia. Jotkin standardiesimerkit feministisessä kirjallisuudessa liittyvät kielen oppimiseen. Evoluutiopsykologiassa esitettiin esimerkiksi pitkään, että kieli kehittyi miehisen metsästysaktiviteetin tuloksena. Koska miehillä oli tarve koordinoida toimintaansa, he kehittivät kielen. Naisten rooli kielen opettajina, kommunikoijina ja erilaisten aktiviteettien koordinoijina puuttui tästä teoriasta pitkään lähes kokonaan (Giere 1996; Longino 1990).

Metodologiassa voi siten nähdä jaon seuraavanlaisiin teorioihin: ensiksi poststrukturalismin ja hermeneutiikan **konstruktivistisiin** teorioihin, joiden mukaan tieteellinen tieto luodaan eli konstruoidaan tiedeyhteisössä, ja toiseksi **objektivistisiin** teorioihin, jotka uskovat tiedon fundamentaaliseen realistiseen perustaan. Näiden välille on esitetty myös välittäviä näkemyksiä (esim. Longino 1990).

Usein uskotaan erheellisesti, että monistinen ontologia välttämättä johtaa monistiseen metodologiseen käsitykseen. Ontologinen monismi ja metodologinen monismi eivät kuitenkaan suoraan edellytä toinen toisiaan. Ontologinen pluralismi ja metodologinen pluralismi eivät nekään implikoi toisiaan. Todellisuutta voidaan hyvin pitää ontologisesti yhdenlaisena tai monimuotoisena mutta silti esittää tieteelliselle tutkimukselle

yhdet yhteiset tai useat erilaiset kriteerit. Esimerkiksi jotkin neutraalin monismin teoriat ymmärtävät todellisuuden systeemien relaatioina: hyvää tieteellistä tutkimusta leimaa siten todellisuuden systeemisen luonteen ymmärtäminen. Kielentutkimuksessa juuri tällainen systeeminen näkökulma metodologiaan on erityisen tuttu ja luonteva: oletetaan kielijärjestelmä ja voidaan tutkia sitä diakronisesta tai synkronisesta näkökulmasta (ks. Johdannon luku 3 tk.). Voidaan tutkia sitä sen osien ja toimintojen kautta. Myös eri kielijärjestelmiä (eri kieliä tai eri järjestelmiä saman kielen sisällä) voidaan verrata toisiinsa. Kielentutkimuksessa kieli ymmärretään usein järjestelmänä tai systeeminä, jota lähestytään jostakin näkökulmasta ja tutkimuskysymyksestä käsin. Näkökulma ja tutkimuskysymys ohjaavat metodin valintaa. Kielentutkimuksessa metodiikkojen moninaisuus ja monimenetelmäisyys on tosiseikka ja elävä osa tieteenalan tutkimusarkea. Kuviossa 11 on esitetty puukuviona tässä luvussa käsitelty metodologinen asemoituminen.




Kuvio 11. Metodologinen asemoituminen kielentutkimuksessa.

## 2.6. Looginen rakenne

Tieteellinen tutkimusasetelma on argumentin tai perustelun rakentamista. Tämä tarkoittaa, että tieteellisellä tutkimuksella on looginen rakenne. 1600-luvulla empiristit alkoivat korostaa yhä voimakkaammin, että tieteessä tutkimusasetelmien tai metodien tulisi rakentaa todellisuuden kanssa sellainen vuorovaikutuksellinen suhde, jonka kuvaamista tieteellisessä julkaisussa voisi käyttää väitteen perusteena. Kuka tahansa, joka olisi saanut riittävän koulutuksen, voisi ymmärtää, miten metodia soveltamalla on päästy pääväitteeseen. Tutkimuksilta alettiin vaatia siis aivan tietynlaista loogista rakennetta. Samalla vakiintui käsitys, että tieteen tehtävä on antaa selityksiä. Selitys eroaa argumentista siinä, että se esittää syyn tapahtuneelle, kun taas argumentti esittää perustelut väitteelle uskoa jokin asia. Esimerkiksi:

> Argumentti: ”Mato on kuollut, koska elonmerkkejä ei havaita.”

> Selitys: ”Mato on kuollut kuivumiseen.”

Selitysten erityinen arvo on siinä, että niitä voi käyttää ennusteina. Jos jonkin tapahtumaketjun alkusyy tunnetaan, voidaan ennustaa tuleva tapahtuma.

> Ennuste: ”Kuivuminen aiheuttaa madolle vakavan terveysriskin ja lopulta kuoleman.”

Valistuksesta alkusysäyksensä saanutta tieteellistä vallankumousta ja sen mukanaan tuomaa teollista vallankumousta ja teknologiakehitystä seuraavien vuosisatojen aikana pidetään usein juuri selittävän tutkimuksen ja pätevien ennusteiden ansiona. Viime vuosisadalla tieteellisen tutkimuksen ajateltiin olevan myös demokratian ja toimivan yhteiskunnan edellytys. Usein nykykäsitys näyttää kuitenkin olevan, että teknologisessa ja taloudellisessa kilpajuoksussa menestyminen vaatii pelkästään selittävää tutkimusta. Voidaan siis edelleen havaita preferenssi selittävään tutkimuksen, jossa selvennetään ilmiön syitä (vrt. kuvaileva tai

ymmärtävä tutkimus (ks. Johdannon luku 2.5. tk.). Ongelma tällaisissa preferensseissä on yleensä se, että ei huomata deskriptiivisen ja syntetisoivan perustutkimuksen olevan usein selittävän tutkimuksen edellytys.

Selittävän tutkimusasetelman rakennetta alettiin pohtia siis jo 1600-luvulla. Esimerkiksi termi *eksperimentti* (koe, koeasetelma) alkoi tällöin esiintyä useamman tieteenfilosofin kirjoituksissa. Aluksi termillä näytettiin tarkoitettavan vain sellaista tutkimusasetelmaa, joka käyttää kokemustietoa. Ensimmäisen eksperimentin julkaisi italialainen lääkäri Francisco Redi (1627–1697). Hän selvitti, pitääkö paikkansa se Aristoteleen (1994) väite, että kärpäset syntyvät spontaanisti maatuvasta aineesta. Redin päämääränä oli esittää perusteita vastakkaiselle näkemykselle. Redin koe perustui siihen, että kahteen ryhmään lasipurkkeja sijoitettiin maatuvia aineita mutta vain toisen ryhmän purkkeihin laitettiin kansi päälle. Tulos osoitti, että vain kannettomiin purkkeihin ilmestyi kärpäsiä, eli väite kärpästen spontaanista syntymisestä ei pitänyt paikkaansa (Lockshin 2007, 142; Konstenius 2014, 40).

John Stuart Mill analysoi käsitteellisesti Redin käyttämät metodologiset periaatteet. Mill järkeili, että kokeellinen tutkimusasetelma perustuu erottavan tekijän periaatteeseen. Erottavan tekijän periaate esittää, että tutkimusasetelma rakennetaan siten, että voidaan seurata kahta samanlaista prosessia, jotka eroavat vain yhden tekijän osalta. Tällöin on mahdollista seurata sitä, onko tämä yksi tekijä se, joka saa aikaiseksi oletetun vaikutuksen. Jos siis halutaan osoittaa, että A on B:n syy ($A \rightarrow B$), on suljettava pois jokainen muu mahdollinen vaikuttava tekijä kontrolloimalla ne. Kontrolloiminen ja poissulkeminen tapahtuvat vertaisryhmän avulla. Kun ryhmillä C1 ja C2 on samat kontrolloidut olosuhteet mutta vain ryhmä C1 käsitellään riippumattomalla muuttujalla A:lla, saadaan selville, onko A:lla vaikutusta B:hen. Vertailemalla ryhmiä käsittelyn jälkeen voidaan havaita, tapahtuiko vaikutus B vain ryhmässä C1. Eksperimentti perustuu näin ollen eliminatiiviseen inferenssiin, eli suljetaan pois vaihtoehtoiset selitykset kontrolloimalla ne eksperimentissä vertaisryhmän avulla (Godfrey-Smith 2003, 212; Konstenius 2014, 40–45).

Eksperimentti on tutkimusasetelma, jolla on seuraava looginen rakenne:

Hypoteesi: A → B (A aiheuttaa B:n)

Kilpailevat hypoteesit: Hn → B (muut tekijät, jotka voivat aiheuttaa B:n)

Ryhmä C1 ja C2 (ryhmä C1, johon vaikutetaan, ja kontrolliryhmä C2)

t1 = alkuhetki ja t2 = päätöshetki (kokeen alkamis- ja päättämishetki).

Seurataan samanaikaisesti kahta eri prosessia. B on A:n seuraus vain, jos B seuraa ajallisesti A:n jälkeen, ja ainoastaan siinä tapauksessa, että ilman A:ta B ei esiinny:

t1 AHn(C1) → t2 C1

t1 Hn(C2) → t2 C2

Mill ja hänen aikalaisensa osoittivat, että tutkimusasetelmilla on usein loogisen argumentin rakenne. Tutkimusasetelma tai tieteellinen metodi ei siten ole mikä tahansa havainnointitilanne tai tapahtuma. Tieteellinen metodi on käytännöllinen prosessi, joka päämäärähakuisesti toteuttaa rakennetta, jonka tieteellisessä argumentaatiossa on määrä vakuuttaa järki. Tutkimusprosessi vuorostaan on se ajallis-paikallinen tapahtuma, jossa tuota rakennetta suoritetaan (ks. Kakkuri-Knuuttila 1998).

Argumentaatiossa sellainen esitys, joka perustuu logiikkaan ja joka sulkee pois muut vaihtoehtoiset hypoteesit, on hyvin vakuuttava. Eksperimentti ja siitä johdettu kausaaliselitys muodostivat siksi pitkään tieteen ihanteen, sillä sen koettiin tarjoavan muita tutkimusasetelmia paremmin sellaista evidenssiä, jonka perusteella voidaan selittää ja ennustaa todellisuuden tapahtumia.

Pitkällä aikavälillä metodologian ja logiikan alan työ kuitenkin osoitti, että tutkimusasetelmilla ja metodeilla yleisesti on parhaimmillaan

sellainen looginen rakenne, joka on pätevää päättelyä suhteessa tutkimuskysymykseen. Kausaaliselitystä etsivä tutkimusasetelma onkin järkevä vain silloin, kun etsimme selittävää tekijää eli syytä. Jos tutkijaa esimerkiksi kiinnostaa subjektien kokemus tai vaikkapa jokin määrällinen kysymys, tulee hänen käyttää toisenlaisia metodeja. Tutkimuskysymys siis ohjaa metodin valintaa ja sen arviointia, onko tutkimusprosessin rakenne looginen vastaus kysymykseen. Jos esimerkiksi halutaan tietää, millä tavoin suomen kielen itsenäiset infinitiivit toimivat affektin ja empatian konstruktioina (esim. *miettiä nyt tällaisia asioita,* ks. Visapää 2013), ei kokeellinen tutkimusasetelma, jossa koetellaan selittävää tekijää, ole kovin mielekäs. Sitä vastoin elokuva, kirja tai verkkokeskustelu voivat tarjota sellaista materiaalia, josta tutkija voi rakentaa aineistoon perustuvan tutkimuksen, jossa tarkastellaan sitä, millä tavoin tunnepitoinen merkitys rakentuu kielen ja kontekstin yhteisvaikutuksessa. Kuten monet kielen ilmiöt affektiivisuus on vuorovaikutuksen läpinäkyvä ilmiö, jonka ilmentymät kielessä ovat systeemisessä vuorovaikutuksessa niin kielisysteemin sisäisten kuin ulkoisten tekijöiden kanssa. Kieli-ilmiöiden erottelu ja analyysi edellyttävät kielentutkijan usein myös vertaavan luokitteluaan ja analyysiaan vastaesimerkkien valossa. Vastaesimerkkien käyttö kielitieteellisessä argumentaatiossa, esimerkiksi ei-affektiivinen kielenkäyttö, voimistaa esitetyn argumentin evidenssiperustaa.

Hyvä tieteellinen argumentaatio, pätevä päättely ja tieteellinen tutkimusasetelma implikoivat toinen toistaan reaalitieteissä. Yleinen näkemys on se, että tutkimuksen kohde ja sen ontologia määräävät metodin perusteita ja sitä, mikä on pätevää päättelyä. Esimerkiksi Jane Austenin tuotantoa ei voi tulkita kokeellisesti, eikä farmakologiassa saa tietoa molekyylien rakenteesta hermeneuttisella tulkinnalla. Toivottavaa olisi, että tutkittaisiin lisää ja seikkaperäisemmin sitä, millä tavoin tieteellisiä argumentteja rakennetaan juuri kielentutkimuksessa. Yleistä on, että argumentaation rakenne vaihtelee metodiikan eri alueilla. Metodiikka jaetaan yleensä neljään osaan:

1. Aineiston keräämisen tekniikat: otantatekniikka ja kokeiden suunnittelu, havainnointi, eksperimentit (koeasetelmat), kyselyt, mittaaminen ja kerääminen.
2. Aineiston kategorisoimisen ja kuvaamisen tekniikat: laadullisen aineiston luokittelu, numeeristen aineistojen muodostaminen, korpusaineistojen annotointi, tunnuslukujen ja kategorioiden muodostaminen.
3. Aineiston analyysin tekniikat: tilastolliset ja matemaattiset menetelmät, taustateoriat laadullisessa jäsennyksessä.
4. Päätelmien tekemisen tekniikat: päätösteoriat, virheteoria, kausaalianalyysi, tilastollisten testien teoria.

Tieteen kehittyessä jokaiseen edellä esitettyyn metodiikan vaiheeseen 1–4 liittyy eri aloilla erilaisia vaatimuksia argumentin rakentamisesta. Esimerkiksi aineistojen keräämiseen liittyy usein erilaisia alakohtaisia perusteluja ja rationaliteetteja sen mukaan, minkä tyyppistä argumenttia ollaan rakentamassa ja millainen on se ekologinen validiteetti (vastaavuus todellisuuden kanssa), johon pyritään. Esimerkiksi yhtä kieltä koskevan korpuslingvistiikan ja monta kieltä rinnakkain tutkivan typologian tavat kerätä ja tasapainottaa aineistoa eroavat toisistaan. Otantatekniikalla, joka on menettelytapa otoksen määrittämiselle jostakin perusjoukosta, pyritään osoittamaan, että tutkimuksen tuloksena oleva väitelause kattaa sen alueen, johon se on yleistetty. Pyrkimys validiteettiin toisin sanoen vaikuttaa kullakin kielentutkimuksen osa-alueella argumentaation rakenteeseen ja logiikkaan.

Sellaisiin tutkimusasetelmakokonaisuuksiin, joissa yhdistyvät implisiittisesti jotkin edellä esitetyt metodiikan osat 1–4, tai erityisiin mittaamisen mahdollistaviin kokonaisuuksiin viitataan toisinaan myös *paradigma*-termillä. Tieteenalalle vakiintunutta paradigmaa monesti toistetaan ilman, että sen perusteita enää pohditaan erikseen tutkimusasetelmissa. Operationaaliset määritelmät takaavat tutkimuksen toistettavuuden, mutta ne ovat luonteeltaan keinotekoisia ja usein käsitteellisesti köyhiä. Tavallisesti jokin tietty operationaalinen määritelmä vakiintuu tieteellisessä kirjallisuudessa tiettyyn tutkimusasetelmaan siten, että tutkijat käyttävät keskenään samaa operationaalista määritelmää

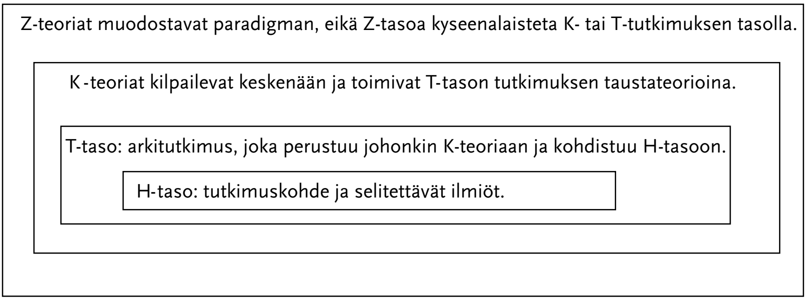
riippuvan ja riippumattoman muuttujan määrittelemiseen paradigman sisällä (Forzano & Gravetter 2009, 74). Kaikki paradigmat perustuvat kuitenkin joukkoon taustaoletuksia, jotka eivät välttämättä aina ole oikeita.

Termi *paradigma* on peräisin Thomas Kuhnilta (1962), joka esitti, että normaalitieteessä tutkimusasetelmien rakentamista ohjaa joukko taustaoletuksia. **Normaalitiede** on vaihe, jossa joukko ilmiöitä on hyväksytty tosiseikoiksi ja suuri osa harjoitetuista tutkimuskäytännöistä kyseenalaistamattomiksi metodeiksi. Näiden tosiseikkojen ja metodien hyväksyminen perustuu taustaoletuksiin, joita ei kuitenkaan välttämättä enää kerta kerralta lausuta tutkimuksen kuluessa vaan tutkimusasetelmaa toistetaan tiedeyhteisön hyväksyessä taustaoletukset. Paradigman vaihdos tai muutos tapahtuu tieteessä silloin, kun tiedeyhteisö tai kilpaileva koulukunta ei hyväksy taustaoletuksia. Kuhnin mukaan standarditutkimuksessa kopioidaan tutkimusasetelma, termit ja osittain tuloksetkin tai usein ainakin tulosten esittämistavat. Standarditieteessä ei kyseenalaisteta paradigmaa eli taustateoriaa ja metodia. Nykyään ajatellaan, että Kuhnin käsitys yhdestä standarditieteestä oli liian yksinkertainen. Tosiasiassa tieteenalan sisällä toimii joukko suuntauksia, joissa toimii keskenään kilpailevia koulukuntia. Koulukuntien välillä voi olla enemmän tai vähemmän yksimielisyyttä hyväksyttyjen paradigmojen laadusta ja määrästä. Tieteellinen tutkimus on siten kirjavampi, heterogeenisempi ja dynaamisempi myös ns. standarditieteen aikana kuin, mitä Kuhn näytti ajattelevan.

Kuhnin oppilas Philip Kitcher (1981) on esittänyt näkemyksen, että tieteellinen selitys q on relaatio, joka propositiolla on selitettävään ilmiöön p, mutta selitys q on hyväksyttävissä vain suhteessa johonkin metateoriaan. Kitcherin määritelmän mukaan metateoria on systematisaatio, joka parhaiten selittää joukon hyväksyttyjä faktoja ilmiöstä p. Tällainen selittävä metateoria on esimerkiksi Newtonin mekaaninen maailmankuva, Einsteinin suhteellisuusteoria tai Darwinin evoluutiooppi. Metateoria selittää joukon aikaisemmin irrallisia havaintoja ja ennustaa tulevia havaintoja, mutta siihen viitataan usein vain implisiittisesti varsinaisissa tutkimusjulkaisuissa. Mitä hyväksytympi on paradigma, sen implisiittisemmäksi sen tietorakenne muuttuu tutkimusjulkaisuissa, sillä tiedeyhteisön konsensus liittää sen automaattisesti

osaksi kollektiivisesti omaksuttua metateoriaa, jonka valossa uusia selityksiä arvioidaan.

Kuviossa 12 selvennetään Kitcherin ajatuksia yhdistävästä selittämisestä (engl. *Explanatory Unification*). Systematisoivat metateoriat, kuviossa 12 Z-teoriat, muodostavat tiettynä ajankohtana tutkimuksen paradigman, jota ei tutkimuksen kuluessa kyseenalaisteta. Z-teoriat muodostavat keskenään mahdollisimman yhtenäisen maailmankuvan, eivätkä ne kilpaile toistensa kanssa. K-teoriat ovat jonkin alan varsinaisia tutkimuksen taustateorioita. Tällaisia ovat kielentutkimuksessa esimerkiksi konseptuaalinen semantiikka, kieliopillistumisteoria tai historiallis-vertaileva menetelmä. K-teoriat kilpailevat keskenään, ja niiden tavoitteena on arkitutkimuksen tasolla selittää mahdollisimman paljon tutkimuskohteen ilmiöitä koherentisti.




Kuvio 12. Kitcherin (1981) malli yhdistävästä selittämisestä.

On syytä panna merkille, että Kitcher käyttää *metateoria*-termiä eri merkityksessä kuin, miten se tämän kokoelman alussa määriteltiin. Edellä määriteltiin metateoria ylätason tutkimusnäkökulmaksi tai tutkimusalaksi. Kitcher vuorostaan käyttää termiä **implisiittisen taustateorian** merkityksessä: Z-tason teoriat ovat mykkiä todeksi hyväksyttyjen taustateorioiden joukkoja, jotka ohjaavat varsinaisten eksplikoitujen teorioiden rakentamista.

Kun tarkastellaan kielentutkimusta kitcheriläisessä käsitteistössä, tieteenalan sisällä voidaan havaita erilaisia metateoreettisia koulukuntia. Huomataan, että on olemassa eriäviä paradigmaattisia näkemyksiä kielen olemuksesta. Esimerkiksi generativistit (formalistit) ja funktionalistit

tekevät tutkimusta eri metateorian sisällä. Metateorian keskeisiä oletuksia (omaksuminen vs. oppiminen, komputationalismi vs. konnektionismi, rationalismi vs. empirismi jne.) ei eksplikoida yhä uudestaan tutkimusjulkaisuissa, vaan ontologiset, tietoteoreettiset ja metodologiset perusteet oletetaan yleensä julkaisuissa tunnetuiksi ja yleisesti hyväksytyiksi kunkin suuntauksen sisällä.

T-tasolla, esimerkiksi semantiikassa tai pragmatiikassa, voi siten työskennellä rinnakkain samankaltaisin metodein eri koulukuntien edustajia. Tämä tarkoittaa sitä, että käytössä voi olla vaikkapa sama havainnointiin perustuva otanta- ja havainnointitekniikka mutta taustateoria K on erilainen; toinen käyttää teoriaa K1 ja toinen teoriaa K2 tutkimuksessaan. Kun tutkimuksessa metodit ovat samankaltaiset ja ainoastaan taustateoriat eroavat toisistaan, tulee olla erityisen tarkka samaan ilmiöön viittaavien termien kanssa, sillä päällepäin termeillä näyttäisi olevan viittauskohteensa ja homonyymisyytensä vuoksi sama merkitys, mutta erilaisten taustateorioidensa vuoksi ne ovat eri tavoin teoriapitoisia (ks. Johdannon luku 2.4. tk.). Toinen termi on K1- ja toinen K2-pitoinen, sillä niiden merkityssisältöön vaikuttaa taustateorian käsitejärjestelmä. Tällaiset termit, joilla on sama muoto mutta jotka ovat eri tavoin K-pitoisia, edustavat taustakäsitejärjestelmiensä perusteella eri käsitteitä. Esimerkiksi *kognitio* viittaa Langackerin kognitiivisessa kieliopissa (1987) erilaiseen ilmiöön kuin Searlen (1992) tai Roschin (1975) tutkimuksissa (ks. Kenttä 2003).

Jotta teorian käsitejärjestelmä pysyisi koherenttina ja loogisena, sitä koskee joukko vaatimuksia (Karihalme 1996). Tällaisia ovat esimerkiksi seuraavat:

1) Yksimerkityksisyys: määritelmät annetaan siten, että jokaista merkitystä käsitejärjestelmässä nimeää tarkkaan vain yksi termi ja tämä termi nimeää vain yhtä ilmiötä.[10]
2) Perusteltavuus: käsite on motivoitu taustateoriassaan.
3) Informatiivisuus: käsitteen olemassaolo kertoo jotakin uutta mallista, jota se kuvaa.
4) Järjestelmällisyys: käsite organisoi koko käsitejärjestelmän tietosisältöä.
5) Rajaavuus: käsitteen ekstension alaa kontrolloidaan teoriassa.

Tieteen tavoitteena pidetään informatiivisia totuuksia, ja oletus on, että yllä olevia ohjeita noudattamalla tähän tavoitteeseen päästään. Oletetaan, että kun yksimerkityksisyyden, perusteltavuuden, informatiivisuuden ja järjestelmällisyyden ihanteita noudatetaan, teorian käsitejärjestelmä on looginen ja termien referentit tunnetaan täsmällisesti. Lopputuloksena on teksti, joka on informatiivista ja totta.

Tieteellistä käsitteenmäärittelyä koskevat normit ja tavoitteet tulee erottaa tosiasiallisesta terminologian tutkimuksesta. Tiedettä ohjaavat pelisäännöt ovat eri asia kuin luonnollista tai tieteellistä kielenkäyttöä tosiasiassa ohjaavat sosiolingvistiset periaatteet. Normatiivisista määritelmistä poiketen tieteellisten termien funktionaaliset määritelmät perustuvat yhteisesti jaetuille kielenkäytön tavoille (ks. esim. Gambier 1987; 1991; Gaudin 1993; 2005; Pearson 1998; Temmerman 2000). Samalla tavoin kuin tieteen pelisäännöt (eli normatiivinen tieteenfilosofia) ja varsinainen tieteentutkimus (esim. deskriptiivinen tieteenhistoria) ovat kaksi eri tutkimusaluetta, niin tieteellisen käsitteenmuodostuksen ja terminologian tutkimus ovat kaksi erillistä tutkimuksen aluetta. Tieteellinen käsitteenmuodostus on normatiivista, sillä kun normeista poiketaan, termeistä tulee häilyviä ja monimerkityksellisiä eikä ole enää välttämättä selvää, mitä kirjoittaja kulloinkin tarkoittaa. Terminologian tutkimus kuvaa yleensä juuri tätä tosiasiallista termien polysemiaan johtavaa kielenkäyttöä. Ei voida esimerkiksi varmasti sanoa, mitä termillä *kognitio* kulloinkin tarkoitetaan, mikäli sitä käytetään tarkoittamaan sekä mieltä että sen sisältöä, sekä tietoista että tiedostamatonta mieltä, sekä yksilömieltä että intersubjektiivista jaettua mieltä tai sekä abstrakteja ideoita että neurokemiallisia prosesseja.

Tieteellisen käsitteenmuodostuksen normatiivisuus tarkoittaa, että tutkijalla on vapaus määritellä käyttämänsä käsitteet siten, että hänen käsitejärjestelmänsä pysyy loogisena. Se tarkoittaa myös, että tutkijalla on velvollisuus päämäärähakuisesti määritellä käsitteensä (Niiniluoto 1980b, 164). Tuo päämäärä on tieteessä tieto tai totuus, ja päästäkseen tähän päämäärään tieteellisen määritelmän on täytettävä ainakin neljä ehtoa:

1. Määritelmän täytyy ilmaista määriteltävän olemus.
2. Määritelmä ei saa olla kehämääritelmä.

3. Määritelmää ei saa lausua negatiivisin ehdoin, jos se voidaan lausua positiivisin kriteerein.
4. Määritelmää ei saa ilmaista kuvaannollisella tai epäselvällä kielellä.

Käsitteiden määritteleminen vaatii omien ja kilpailevien koulukuntien taustateorioiden tuntemusta. Tutkijan omat käsitteet määritellään suhteessa omaan käsitejärjestelmään sekä kilpailevien koulukuntien käsitteisiin. Määritelmässä kerrotaan selkeästi, millä tavoin oma käsite eroaa muista samaa ilmiötä nimeävistä käsitteistä. Whiteheadin (1929) mukaan teoria on sisäisesti looginen ja sen käsitejärjestelmä koherentti, mikäli seuraavat seikat toteutuvat:

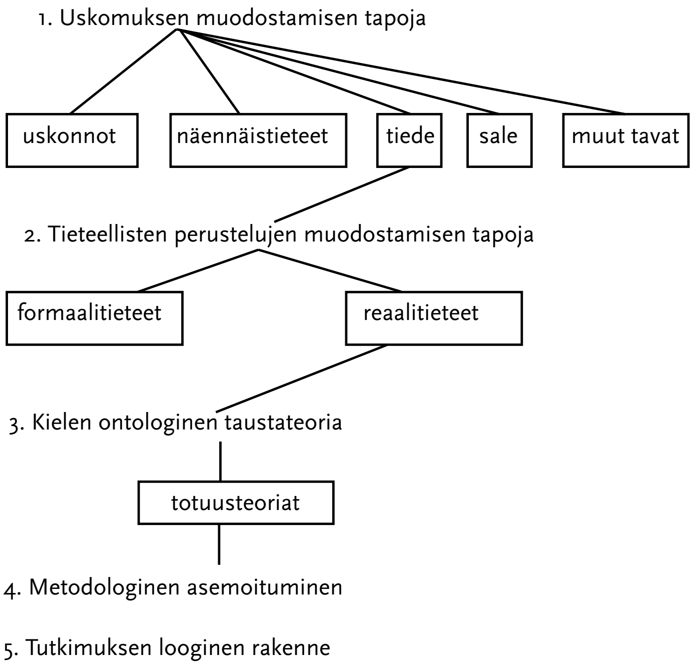
1. Lauseet eivät ole keskenään ristiriitaisia, vaan ne ovat loogisesti konsistentteja.
2. Eri väitteistä seuraavat päätelmät (esim. miten metodeja sovelletaan) eivät ole keskenään ristiriidassa.
3. Väitteet siitä, mitä maailmassa on tai mikä on maailman perimmäinen rakenne, eivät ole keskenään ristiriidassa (onttiset ja ontologiset kysymykset, ks. Johdannon luku 2.4. tk.).
4. Käsitteiden intensio ja ekstensio pysyvät teoriassa konsistentteina.

Tieteellinen käsitteen- ja teorianmuodostus on tavoitteellista toimintaa, jota jokainen tutkija ihanteellisesti harjoittaa tunnollisesti. Tieteellinen käsitteenmuodostus on tieteen ihanne, sillä se johtaa tieteellisesti pätevään tietoon. Tosiasia on, että mitä laajemmassa käytössä tieteellinen termi on, sitä prototyyppisempi on käsitteen rakenne: tietty perusmerkitys yleensä säilyy kategorian ytimessä. Tieteessä käsitteiden merkitysneuvottelut ovat kuitenkin moninaiset. Tieteellisten termien tiukka standardisointi on siten harvoin mahdollista, jos edes toivottavaa. Tieteellisen tutkimuksen tavoite on kuvata, löytää ja selittää uusia asioita. Uudet ilmiöt johtavat usein uusiin määritelmiin ja käsitteisiin. Tiede edistyy, ja se tarkoittaa sitä, että myös määritelmät ja merkitykset kehittyvät.

Jopa yhdenkin tekstin sisällä kirjoittajan käyttämien käsitteiden määritelmien kartoittaminen ja ymmärtäminen voivat olla suuren analyyttisen työn tulosta. Esimerkiksi Wittgensteinin *kielipeli*-käsitteestä on

julkaistu vuodesta 1956 lähtien lukuisia tutkimuksia, eikä tunnu olevan odotettavissa, että tahti hidastuisi. Terminologian tutkimus osoittaa, että tieteelliset käsitteet ovat kiinteä osa taustakäsitejärjestelmäänsä ja teoriaa, eikä atomaaristen standardimääritelmien antaminen ole kovin realististа. Se myös osoittaa, että tieteellisen käsitteenmuodostuksen normien noudattamien on tieteessä erittäin tärkeää. Mitä suurempi on vaara, että tieteellinen kieli muuttuu epätarkaksi, sitä suurempi on tutkijan velvollisuus määrittää käsitteensä.

Kuviosta 13 huomataan, että argumentaation rakenne seuraa metodologista asemoitumista. Metodi on tutkimuksen logiikkaa, joka muodostaa osan väitteen perusteluista ja siten oleellisella tavalla määrittää tavoiteltavan tieteellisen argumentin rakennetta. Tieteellistä argumentaatiota, käsitteen- ja teorianmuodostusta koskee joukko yleisiä periaatteita, joita on myös käsitelty tässä luvussa.




Kuvio 13. Metodi on looginen rakenne.

## 2.7. Etiikka

Metodit ovat tieteen sääntöjä, ja niiden soveltamiseen liittyy joukko eettisiä normeja. Tämä tarkoittaa sitä, että metodin soveltamisen tapaa määrittävät sellaiset oikeaa ja väärää koskevat kysymykset, joihin tiedeyhteisö etsii kollektiivisesti vastauksia. Tiedeyhteisön antamat vastaukset moraalifilosofisiin kysymyksiin ovat normatiivisia, eli ne sitovat ja velvoittavat yksittäistä tutkijaa. Mikäli tieteen tutkimuseettisiä normeja ei tutkimuksessa noudateta, tutkimus ei täytä tieteellisyyden kriteerejä.

**Tutkimusetiikka** tarkoittaa kaikkia niitä oikeaan ja väärään liittyviä kysymyksiä, jotka suoraan kohdistuvat metodiin ja esimerkiksi datan säilytykseen, anonymisointiin, henkilöiden yksityisyyteen ja turvallisuuteen jne. **Tieteen etiikka** liittyy vuorostaan laajemmin tutkijan, tutkimuksen ja tieteellisen tiedon rooliin yhteisössä. Tutkimuskysymys, käsitteistö ja tutkimuksen konteksti kuuluvat tieteen etiikkaan. Tutkimustulosten tulkinta ja julkaisukonteksti ovat niin ikään ns. laajan vastuun aluetta (ks. Johdannon luvut 1 ja 4 tk.).

Tutkimuksen etiikka alkaa tutkimusongelman eettisestä muotoilusta. Kysymyksen muotoilu johtaa tutkimusasetelman arkkitehtuuriin ja niihin alakysymyksiin, joihin tutkimuksen kuluessa vastataan. Jos tutkimus kohdistuu ihmisryhmiin tai sellaisiin ihmisryhmän ominaisuuksiin tai kulttuureihin, joiden muotoutumiseen ryhmän nykyjäsenten voi katsoa olevan osallisia, Creswell (2003, 63) ehdottaa pilottitutkimusta, jotta tutkimusongelma ja tutkimusasetelmassa käytetty käsitteistö muotoiltaisiin heti alussa siten, etteivät ne marginalisoisi tai heikentäisi ryhmän etuja. Tutkimuksen tulokset on julkaistava totuudenmukaisesti, ja siksi on oleellista, että tutkimusasetelma alkuaan on neutraali ja ottaa huomioon osallistujien oikeudet.

Jotta yhteiskunta voisi ottaa kantaa siihen, onko tutkimus yhteiskunnan tavoitteiden ja arvojen mukaista, tutkimuseettisiin neuvottelukuntiin ja toimikuntiin (esimerkiksi Suomen TENK, ks. https://www.tenk.fi/) otetaan tieteellisten jäsenten lisäksi maallikkojäseniä. Vuodesta 2009 lähtien vaatimus ihmiseen kohdistuvan tutkimuksen eettisestä ennakkoarvioinnista on koskettanut Suomessa myös kielentutkimusta. Tämä tarkoittaa sitä, että kielentutkimuksen eettisiä huolenaiheita ei voi

määritellä vain tieteenalan sisältä käsin vaan kielentutkimusta on tarkasteltava myös vallitsevien yleisten ohjeiden, yhteiskunnan keskeisten arvojärjestelmien ja koodien näkökulmasta. Kielentutkijan on muun muassa pidettävä huolta siitä, että hänen käyttämänsä metodologinen käsitteistö vastaa standardimalleja.

Tutkijan on esimerkiksi Tutkimuseettisen neuvottelukunnan (TENK) vaatimukset täyttääkseen kyettävä erottamaan erilaiset metodit ja metodien erityispiirteet toisistaan sellaisin termein, joilla yleiset tutkimuseettiset ohjeet on esitetty. Esimerkiksi *eksperimentti*-termin käyttö edellyttää, että tutkija on käyttänyt kontrolliryhmää vaihtoehtoisten hypoteesien poissulkemiseksi. Erilaisiin tutkimustyyppeihin liittyy erilaisia eettisiä huolenaiheita, joista tutkijan on oltava tietoinen tutkimusasetelmaa rakentaessaan tai ennakkoarviointihakemusta tehdessään. TENK ohjeistaa myös humanisteja metodien totuudesta: vilppi, väärennös ja vääristely ovat rangaistavia. Huolellisuus on enemmän kuin hyve: se on vaatimus.[11]

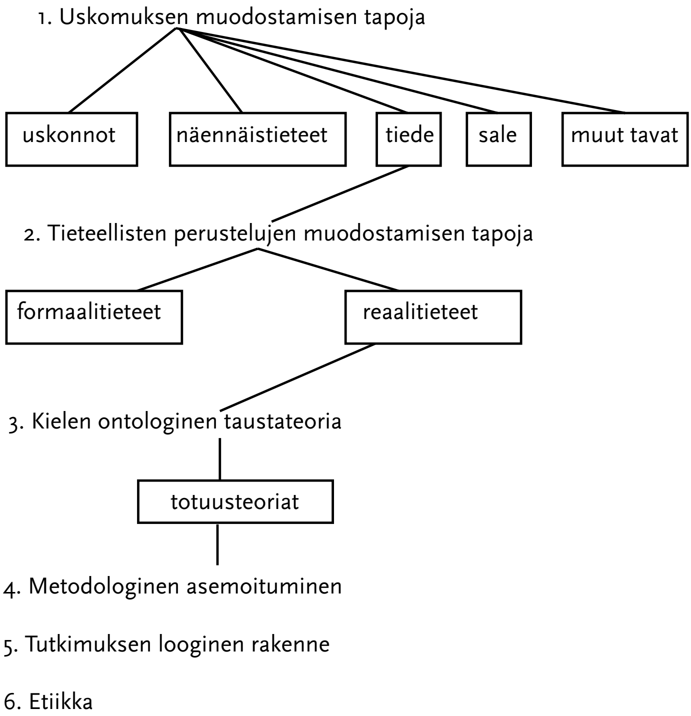
Tieteellinen tutkimus on nykykäsitysten mukaan suoritettava siten, ettei siitä koidu vahinkoa tutkittavalle. Potentiaalisia vahinkoalueita ovat fyysinen ja psyykkinen hyvinvointi, yksityisyys ja henkilön autonomisuus. Kieliaineiston hankkimistapoja tulee kuitenkin reilusti problematisoida suhteessa muiden alojen käytäntöihin ja riskimääritelmiin. Linguistic Society of America (LSA) on esimerkiksi korostanut sitä, että kieliaineiston hankinta on lähes aina luonteeltaan minimiriskistä. Vaikka kenttätutkimuksessa käytettäisiinkin kielenoppaiden pohdinnoista termiä *ajatuskoe*, tätä ”koetta” ei ole mielekästä vain metodin nimen perusteella asettaa samanlaisten sääntöjen ja rajoitusten kohteeksi kuin esimerkiksi biolääketieteellisessä tutkimuksessa käytettyjä aineistoja (kudosnäytteitä, potilasrekistereitä jne.).

Kielentutkija voi tieteellisyyttä korostaakseen haluta käyttää erilaisia *koe-* tai *eksperimentti*-termejä, mutta on tärkeä huomata, että niiden käytön tulisi kuitenkin vastata eksperimenttimetodin tosiasiallisia määritelmiä. Erityisesti kuitenkin etiikkaa pohdittaessa kieliaineiston vahinko- ja riskikriteerejä tulisi problematisoida kielentutkimuksen omilla ehdoilla (Konstenius 2014, 66–70). Muiden alojen terminologiset ratkaisut eivät siis välttämättä ole suoraan sovellettavissa kielentutkimukseen. Tietoisesti tutkijalle julkaistavaksi annettu informaatio alkujaan jo julkisesta

kielitiedosta ei täytä salaamiseen, turvaperiaatteisiin ja kodifikaatioon käytettäviä eettisiä ennakkoehtoja. Myöskään kielentutkimuksen aineistosta johdetut päätelmät eivät LSA:n mukaan aiheuta kieliryhmän jäsenille henkilökohtaista henkistä tai fyysistä vahinkoa. Tämä johtuu siitä, että kielentutkimuksen aineisto päätelmien kohteena ei yleensä ole henkilökohtaisesti ketään yhtä ihmistä koskevaa. Kielitieto on yleensä yhteisön julkista tietoa[12] ja harvoin yksilön näkökulmasta ongelmallista. Sitä vastoin politiikka, uskonto ja etnisyys ovat empiirisen kielidatan keräämisen eettisten valintojen konteksteja, joista tutkijan tulee olla tietoinen (Konstenius 2014, 66–80).

Suomessa TENK (2012) on päättänyt, ettei tutkimuseettistä ennakkoarviointia tarvita, kun kyseessä on normaalipopulaatio. Nuoriltakin saa kysyä, jos tutkimus suoritetaan osana koulun normaalitoimintaa. Sellaiset tutkimukset, joissa ei kerätä yksilöityjä tunnistetietoja, voidaan suorittaa ilman huoltajan lupaa. Muussa tapauksessa tarvitaan ennakkoarviointia. Kielentutkimuksessa tutkimuksen kohteena eivät pääsääntöisesti ole yksittäisen ihmisen vaan kollektiivisen kielen ominaisuudet. Siksi yksityisten ihmisten anonymiteetti voidaan laajoissa aineistoissa varsin helposti taata. McEnery ja Hardie (2011, 66) kuitenkin esittävät, että suojaus on tehtävä aina riippumatta siitä, onko materiaali yksityisluonteista, sillä koskaan ei voi tietää, mitä tulevaisuus tuo tullessaan. Tutkimusetiikkaa käsitellään tarkemmin johdannon luvussa 4.

Kuviossa 14 etiikka on viimeinen niistä ehdoista, jotka kielentutkimuksen on täytettävä, jotta voitaisiin puhua tieteellisestä metodista. Tieteellä on useita erilaisia menettelysääntöjä jokaisessa kuudessa eri tieteellisen metodin ehdossa, kuten tässä luvussa on käynyt ilmi. Viimeisin ja ehkä oleellisin näistä on itse vaatimus noudattaa näitä sääntöjä vilpittömällä rehellisyydellä silloinkin, kun se menee tutkijan omien etujen edelle. Kielentutkimus, kuten muukin tutkimus, kuuluu yhteisen edun piiriin, eli viime kädessä tieteellisestä tiedosta tulisi olla yhteiskunnan jäsenille hyötyä eikä haittaa. Sekä tutkimuseettinen arkipäivän etiikka että laajan vastuun etiikka (ks. Johdannon luku 4 tk.) ovat osa tätä koodistoa, jonka tavoite on taata, että tiede todella tuottaisi niin tutkimuksen kuluessa kuin sen lopullisessa muodossa, tieteellisenä tietona, enemmän hyvää kuin pahaa.




Kuvio 14. Etiikka kuuluu tutkimuksenteon käytäntöön.

## VIITTEET

1 Nominalismin mukaan on olemassa vain reaalisia olioita, ts. teoreettisilla käsitteillä ei ole vastinetta reaalimaailmassa. Nominalismi on siis ontologinen kanta (mitä maailmassa on). Samoin instrumentalismi kieltää teoreettiset oliot ja pitää niitä tai niitä nimeäviä termejä vain enemmän tai vähemmän onnistuneina välineinä tieteellisessä käsitteenmuodostuksessa. Instrumentalismi on tieteenfilosofinen kanta. (Ks. Tieteen termipankki s.v. nominalismi, instrumentalismi.)

2 *Tietoteoria* ja *epistemologia* ovat synonyymejä (ks. Lammenranta 1993).

3 Toisenlaisia jakoja ovat esimerkiksi luonnontieteet vs. ihmistieteet ja selittävät tieteet vs. ymmärtävät tieteet (ks. esim. Niiniluoto 1980).

4 Aiheesta keskustelivat *Virittäjässä* 1972–1975 Esa Itkonen, Esa Saarinen, Bengt-Olof Qvarnström sekä Oiva Ketonen.

5 *A priori* tarkoittaa ennen kokemusta ja *a posteriori* kokemuksen jälkeen. Matemaattiset teoriat ovat analyyttisia a priori, mutta teoreettisen fysiikan matemaattiset teoriat sitä vastoin saavat vahvistusta tai kumoutuvat uuden evidenssin myötä. Esimerkiksi Higgsin bosoni vahvisti hiukkasfysiikan standardimallin. *A priori* on siis totuusteoreettinen termi, joka viittaa proposition totuuden joko kumoutuvaan tai pysyvään luonteeseen. 2 + 2 = 4 on *a priori* (tai apriorisesti) tosi. Sen sijaan Matin tieto siitä, että hänen nimensä on Matti, ei ole *a priori* totta, vaikka Matilla ei ole tarvetta katsoa passiaan tietääkseen, mikä hänen nimensä on.

6 ”If they are known to be true, they must be unfalsifiable, which means that they are non-empirical and thus in some sense a priori” (Itkonen 1997, 8–9).

7 Postpositivismissa ajateltiin vuosisadan alussa, että välittömästi havaittavaa nimeävät termit olisivat ns. semanttisia partikulaareja ja siten ongelmattomia. Filosofinen tutkimus kuitenkin osoitti, että havainto on teoriapitoista (ks. Johdannon luku 3 tk.).

8 Ilmaisu ”kieltää lause” on käytössä logiikassa ja tarkoittaa, että jos A on tosi, niin ei-A ei samanaikaisesti voi olla tosi. Kysymys siitä, mitä tällaiset ristiriitaisuudet teorian sisällä tai teorian ja aineiston välillä tarkoittavat, on yksi tieteenfilosofisen tutkimuksen kohde. Naiivi induktivismi ja naiivi falsifikationismi esittivät, että yksi kieltolause kaataa koko teorian tai hypoteesin. Kehittyneemmät tieteenfilosofiset kehykset yleensä tutkivat kieltolauseita, niiden suhteita ja muita merkitysrakenteita kokonaisuuden näkökulmasta.

9 ”A theory is falsifiable [– –] if there exists at least one non-empty class of homotypic basic statements which are forbidden by it, that is, if the class of potential falsifiers is not empty” (Popper 2002 [1934], 95).

10 Terminologian teoriassa perinteistä yksimerkityksisyyden vaatimusta on myös kritisoitu (ks. esim. Pearson 1998, 7; Temmerman 2000, 219).

11 Ks. tarkemmin http://www.tenk.fi/.

12 Poikkeuksen muodostavat mahdollisesti idiolektit eli yksilökielet sekä sellainen vainotun tai syrjityn kieliryhmän kielitieto, jolla voi olla poliittisia seurauksia (ks. Johdannon luku 4 tk.; Konstenius 2014).

## KIRJALLISUUS

Achterhuis, Herman Johan. 2001. *American Philosophy of Technology: The Empirical Turn*. Bloomington (IN): Indiana University Press.

Allwood, Jens, Andersson, Lars-Gunnar & Dahl, Östen. 1980 [1971]. *Logiikka ja kieli*. [Logik för lingvister]. Suomentanut Paavo Siro [englanninkielisestä laitoksesta *Logic in linguistics* 1977. Cambridge: Cambridge University Press]. Helsinki: Gaudeamus.

Aristoteles. 1994. Toinen analytiikka. [Analytika hystera]. Julkaisussa: Knuuttila, Simo, Niiniluoto, Ilkka, Thesloff, Holger (toim.) *Teokset 1. Kategoriat. Tulkinnasta. Ensimmäinen analytiikka. Toinen analytiikka*. Suomentaneet Lauri Carlson, Simo Knuuttila ja Juha Sihvola. Helsinki: Gaudeamus, 217–242.

Bacon, Francis. 2000 [1620]. *The New Organon*. Cambridge: Cambridge University Press.

Carnap, Rudolf. 1934. *The Unity of Science*. London: Kegan Paul, Trench, Trubner & Co.

Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford: Clarendon.

Chalmers, David. 1996. Facing up to the problem of consciousness. Julkaisussa: Hameroff, Stuart R., Kaszniak, Alfred W. & Scott, Alwyn C. (toim.) *Toward a Science of Consciousness: the First Tucson Discussions and Debates*. Cambridge (MA): MIT Press, 5–28.

Chalmers, David. 2006. Strong and weak emergence. Julkaisussa: Clayton, P. & Davies, P. (toim.) *The Re-emergence of Emergence*: *The Emergentist Hypothesis From Science to Religion*. Oxford: Oxford University Press, 244–255.

Chalmers, David, Manley, David & Wasserman, Ryan (toim.). 2009. *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press.

Couvalis, George. 1997. *The Philosophy of Science: Science and Objectivity*. Thousand Oaks (CA): Sage.

Creswell, John. W. 2003. *Research Design: Qualitative and Quantitative Approaches and Mixed Methods Approaches*. London: Sage.

Croft, William. 2000. *Explaining Language Change: An Evolutionary Approach*. New York (NY): Pearson Education.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Dahl, Östen. 1975. Is linguistics empirical? A critique of Itkonen's Linguistics and Metascience. *Gothenburgs Papers on Theoretical Linguistics* 29. Göteborg: Department of Linguistics, University of Göteborg.

Dennett, Daniel C. 1999. *Tietoisuuden selitys*. [*Consciousness Explained*]. Suomentanut Tiina Kartano. Helsinki: Art House.

Dilthey, Wilhelm 1982 [1860]. *Selected Works, Volume 1. Introduction to Human Sciences*. Toimittaneet Makkreel, Rudolf A. & Firthjof, Rodi. Princeton (NJ): Princeton University Press.

Elga, Adam. 2007. Isolation and folk physics. Julkaisussa: Price, H. & Corry, R. (toim.) *Causation, Physics, and the Constitution of Reality*: *Russell's Republic Revisited*. Oxford: Oxford University Press, 153–173.

Field, Hartry. 2003. Causation in a physical world. Julkaisussa: Loux, M. & Zimmerman, D. (toim.) *Oxford Handbook of Metaphysics*. Oxford: Oxford University Press, 435–460.

Forzano, Lori-Ann B. & Gravetter, Frederick J. 2009. *Research Methods for the Behavioral Sciences*. Belmont (CA): Wadsworth.

Friston, Karl J. & Frith, Christopher D. 2015. Active inference, communication and hermeneutics. *Cortex* 68, 129–143.

Gabriel, Markus. 2015. *Why the World Does Not Exist*. [*Warum es die Welt nicht gibt*]. Kääntänyt Gregory Moss. New Jersey: John Wiley & Sons.

Gadamer, Hans-Georg. 1976. *Philosophical Hermeneutics*. Kääntänyt David Linge. Berkeley: University of California Press.

Gallagher, Shaun & Allen, Micah. 2016. Active inference, enactivism and the hermeneutics of social cognition. *Synthese*, 1–22.

Gambier, Yves. 1987. Problèmes terminologiques des pluies acides: pour une socio-terminologie. *Méta* 32, 314–320. DOI: 10.7202/002791ar.

Gambier, Yves. 1991. Travail et vocabulaire spécialisés: prolégomènes à une socio-terminologie. *Méta* 36, 8–15. DOI: 10.7202/002795ar.

Gaudin, François. 1993. *Socioterminologie: des problemes semantiques aux pratiques institutionnelles*. Rouen: Publications de l'Université de Rouen.
Gaudin, François. 2005. La socioterminologie. *Langages* 157, 80–92. DOI: 10.3917/lang.157.0081.
Giere, Ronald N. 1996. The feminism question in the philosophy of science. Julkaisussa: Nelson, L. H. & Nelson, J. (toim) *Feminism, Science, and the Philosophy of Science*. Synthese Library vol. 256. Dordrecht: Springer, 3–15.
Godfrey-Smith, Peter. 2003. *Theory and Reality. An Introduction to the Philosophy of Science*. Chicago (IL): University of Chicago Press.
Habermas, Jürgen. 1968. *Ciencia y técnica como ideología*. Madrid: Tecnos.
Habermas, Jürgen. 1970. *Knowledge and Human Interests*. New Jersey: John Wiley & Sons.
Hacking, Ian. 2001. *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge University Press.
Hansson, Bengt. 1974. *Explanations – or What*? Stanford (CA): Stanford University.
Hempel, Carl G. 1958. *The Theoretician's Dilemma: a Study in the Logic of Theory Construction*. Minnesota Studies in the Philosophy of Science. Minnesota: University of Minnesota Press.
Hempel, Carl G. & Oppenheim, Paul. 1948. Studies in the logic of explanation. *Philosophy of Science* 15:2, 135–175.
Hume, David. 2016 [1748]. An enquiry concerning human understanding. Julkaisussa: Cahn, Steven M. (toim.) *Seven Masterpieces of Philosophy*. New York (NY): Routledge, 191–284.
Itkonen, Esa. 1975. Vastine edelliseen. *Virittäjä* 79:4, 397–403.
Itkonen, Esa. 1997. The social ontology of linguistic meaning. *Yearbook of the Linguistic Association of Finland* 10, 49–80.
Kaisto, Janne. 2005. *Lainoppi ja oikeusteoria. Oikeusteorian perusteita aineellisen varallisuusoikeuden näkökulmasta*. Helsinki: Edita.
Kant, Immanuel. 1981 [1781]. *Critique of Pure Reason*. [*Kritik der Reinen Vernunft*]. London: Macmillan.
Kakkuri-Knuuttila, Marja-Liisa (toim.). 1998. *Argumentti ja kritiikki. Lukemisen keskustelun ja vaikuttamisen taidot*. Helsinki: Gaudeamus.
Karihalme, Oili. 1996. *Muotoilun teoriasanaston termistyminen*. Acta Wasaensia No 51, Kielitiede 10. Vaasa: Vaasan yliopisto.
Kenttä, Reetta-Leena. 2003. Langackerin kognitiivisen kieliopin metateoreettinen analyysi: Kognitio-käsite ja tieteellinen argumentaatio kognitiivisessa kielitieteessä. Yleisen kielitieteen pro gradu -tutkielma. Helsingin yliopisto, Yleisen kielitieteen laitos.
Ketonen, Oiva. 1975. Kielioppi ja kokemus. *Virittäjä* 79:2, 195–198.
Ketonen, Oiva. 1976. *Se pyörii sittenkin: tieteenfilosofian peruskysymyksiä*. Porvoo: WSOY.
Kitcher, Philip. 1981. Explanatory unification. *Philosophy of Science* 48:4, 507–531.
Kitcher, Philip. 1983. *Abusing Science: the Case Against Creationism*. Cambridge (MA): MIT Press.
Konstenius, Reetta A. 2014. *Empiria, eksperimentti ja etiikka. Kielitieteen metateoriaa*. Helsinki: Helsingin yliopisto.

Korhonen, Mikko. 1986. *Finno-Ugrian language studies in Finland 1828–1918*. Helsinki: Societas Scientiarum Fennica.
Kuhn, Thomas S. 1962. *The Structure of Scientific Revolution*. Chicago (IL): University of Chicago Press.
Kuhn, Thomas S. 1971. Concepts of cause in the development of physics. Julkaisussa: Kuhn, T. *Essential Tension*: *Selected Studies in Scientific Tradition and Change*. Chicago (IL): The University of Chicago Press, 21–30.
Lakoff, George & Johnson, Mark. 1999. *Philosophy in the Flesh*. New York (NY): Basic books.
Lammenranta, Markus. 1993. *Tietoteoria*. Helsinki: Gaudeamus.
Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar: Theoretical prerequisites*. Stanford (CA): Stanford University Press.
Laplace, Pierre Simon. 1951 [1814]. *A Philosophical Essay on Probabilities*. [*Essai philosophique sur les probabilités*]. Kääntänyt Truscott, F. W. & Emory, F. L. New York (NY): Dover Publications.
Latham, Noa. 1987. Singular causal statements and strict deterministic laws. *Pacific Philosophical Quarterly* 68, 29–43.
Lewes, George Henry. 1875. *Problems of Life and Mind*. Vol. 2. London: Kegan Paul, Trench, Turbner & Co.
Lockshin, Richard. 2007. *The Joy of Science: An Examination of how Scientists Ask and Answer Questions Using the Story of Evolution as a Paradigm*. Dordrecht: Springer.
Loewer, Barry. 2001. Review of J. Kim, Mind in a Physical World. *Journal of Philosophy* 98, 315–324.
Longino, Helen. E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton (NJ): Princeton University Press.
Lyas, Colin (toim). 1971. *Philosophy and Linguistics.* New York (NY): St Martin's Press.
McEnery, Tony & Hardie, Andrew. 2011. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
Merton, Robert K. 1973 [1942]. The normative structure of science. Julkaisussa: Merton, Robert K. (toim.) *The Sociology of Science. Theoretical and Empirical Investigations*. Chicago (IL): University of Chicago Press, 223–280.
Mill, John S. 1906 [1875]. *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation*, Vol. 1. London: Longman.
Määttä, Urho. 2000. Emergentistinen kielitiede. Julkaisussa: Määttä, Urho, Nieminen, Tommi & Pälli, Pekka (toim.) *Emergenssin kielelliset kasvot*. Folia Fennistica & Linguistica 24. Tampere: Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos, 17–39.
Niiniluoto, Ilkka. 1980a. *Is Science Progressive?* Dordrech: Kluwer.
Niiniluoto, Ilkka. 1980b. *Johdatus tieteenfilosofiaan. Käsitteen- ja teorianmuodostus*. Helsinki: Otava.
Niiniluoto Ilkka. 1983. *Tieteellinen päättely ja selittäminen*. Helsinki: Otava.
Niiniluoto, Ilkka. 1990. *Maailma, minä ja kulttuuri: emergentin materialismin näkökulma*. Helsinki: Otava.
Niiniluoto, Ilkka. 1999. *Critical Scientific Realism*. Oxford: Oxford University Press.

Niiniluoto, Ilkka. 2001. Julkisuusperiaate ja tutkimustulosten salailu. *Tieteessä tapahtuu* 19:2.
O'Connor, Timothy & Hong, Yu Wong. 2005. The metaphysics of emergence. *Noûs* 39:4, 658–678.
Pearson, Jennifer. 1998. *Terms in Context*. Amsterdam: John Benjamins. DOI: 10.1075/scl.1.
Popper, Karl. 2002 [1934]. *The Logic of Scientific Discovery*. London: Routledge.
Putnam, Hilary. 1977. *Reason, Truth and History*. Cambridge: Cambridge University Press.
Quine, Willard Van Orman. 1948. On what there is. *The Review of Metaphysics*, 21–38.
Raatikainen, Panu. 2004. *Ihmistieteet ja filosofia*. Helsinki: Gaudeamus.
Raatikainen, Panu. 2007. Mentaalinen kausaatio. Julkaisussa: Gylling, Heta, Niiniluoto, Ilkka & Vilkko, Risto (toim.) *Syy*. Helsinki: Gaudeamus, 227–237.
Redhead, Michael. 1990. Explanation. Julkaisussa: D. Knowles (toim.) *Explanation and Its Limits*. Cambridge: Cambridge University Press, 135–54.
Reichenbach, Hans. 1938. *Experience and Prediction*. Chicago (IL): University of Chicago Press.
Rosch, Eleanor. 1973. Natural categories. *Cognitive Psychology* 4:3, 328–350.
Rosch, Eleanor. 1975. Cognitive representations of semantic categories. *Journal of Experimental Psychology, General* 104:3, 192–233.
Saarinen, Esa. 1975. Kieliopin testattavuus. *Virittäjä* 79:4, 389–398.
Salmon, Wesley C. 1971. *Causality and Explanation*. Oxford: Oxford University Press.
Scheuermann, Leif & Kroeze, Jan H. Hendrik. 2017. *Digital Humanities and Information Systems: Innovating Two Research Traditions*. [verkkoaineisto]. [viitattu 4.2.2017]. AIS Electronic Library. Saatavissa: http://aisel.aisnet.org/amcis2017/PhilosophyIS/Presentations/3/.
Searle, John R. 1992. *The Rediscovery of the Mind*. London: MIT Press.
Sellars, Wilfrid. 1963. *Science, Perception and Reality*. London: Routledge.
Suppe, Frederick. 1974. *The Structure of Scientific Theories*. Illinois: University of Illinois Press.
Taylor, Charles. 1971. Interpretation and the science of man. *The Review of Metaphysics* 25:1, 3–51.
Temmerman, Rita. 2000. *Towards New Ways of Terminology Description: The Sociocognitive Approach*. Amsterdam: John Benjamins. DOI: 10.1075/tlrp.3.
TENK 2012. Hyvä tieteellinen käytäntö ja sen loukkausepäilyjen käsitteleminen Suomessa. Tutkimuseettisen neuvottelukunnan ohje 2012. Helsinki: Tutkimuseettinen neuvottelukunta. [verkkoaineisto]. [viitattu 13.8.2018]. Saatavissa: http://www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf.
Thagard, Paul R. 1978. Why astrology is a pseudoscience. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1978:1, 223–234.
Thomasson, Amie. 2007. *Ordinary Objects*. New York (NY): Oxford University Press.
Tieteen termipankki. 2019. [verkkoaineisto]. [viitattu 28.2.2019]. Saatavissa: https://tieteentermipankki.fi.
Tuomela, Raimo. 1973. *Theoretical Concepts*. Vienna: Springer-Verlag.
Visapää, Laura. 2013. Itsenäiset infinitiivit affektin ja empatian konstruktioina. *Virittäjä* 117:4, 524–550.

Weber, Max. 1949. *The Methodology of the Social Sciences*. Kääntänyt ja toimittanut Edward A. Shils & Henry A. Finch. New York (NY): Free Press.
Whitehead, Alfred N. 1929. *The Function of Reason*. Princeton (NJ): Princeton University Press.
Wittgenstein, Ludwig. 2009 [1953]. *Philosophical Investigations*. [*Philosophische Untersuchungen*]. Kääntänyt G. E. Anscombe. Malden (MA): Wiley-Blackwell.
Wright, Georg Henrik von. 1999 [1971]. *Tieto ja ymmärrys*. Helsinki: Otava.

# 3. Monimenetelmäisen kielentutkimuksen metateoreettisia lähtökohtia

*Markus Hamunen*
https://orcid.org/0000-0003-2786-6392

*Reetta Konstenius*

Johdannon luvussa 2 käsiteltiin tieteellisen menetelmän yleisiä ehtoja. Tämä luku jatkaa siitä kielentutkimuksen monimenetelmäisyyteen ja sen metateoreettisiin perusteisiin. Yleisesti monimenetelmäisyys on sitä, että tieteenalalla käytetään useita eri metodeja ja näkökulmia tutkimuskohteen erilaisten ominaisuuksien selvittelyyn. Tämä on luonnollista kielen kaltaisen monimutkaisen ilmiön tutkimisessa. Käsitteellisenä erontekona on hyvä tiedostaa, että tässä luvussa avattava yksittäisen tieteenalan **monimenetelmäisyys** on eri asia kuin yksittäistutkimuksen **monimenetelmäinen tutkimusasetelma** (engl. *mixed methods*). Se on myös eri asia kuin **metodologinen pluralismi**, joka on normatiivinen tieteenfilosofinen kanta sille, että tieteessä tarvitaan erilaisia malleja, selityksiä tai metodeja. Tämän luvun päämääränä ei ole tarjota perusteellista kuvaa tieteenfilosofiasta (nyky)kielentutkimuksessa, vaan tarkoituksena on esittää yleisiä aihepiiriin kuuluvia teemoja. Tämän luvun tavoitteena on selkeyttää kielitieteelliseen tutkimusasetelmaan, metodiin ja teorianmuodostukseen vaikuttavia taustaoletuksia. Erilaiset kielen olemusta ja tiedonhankintaa koskevat uskomusjärjestelmät ohjaavat välttämättä sitä, miten tutkimus asemoituu suhteessa kohteeseensa ja miten tutkimuskäsitteistö valitaan ja miten sitä sovelletaan. Monimenetelmäistä tutkimusalaa koskevat myös jotkin sellaiset kielentutkimuksen luonteen taustaoletukset, joita on mielekästä selventää johdannon tässä luvussa.

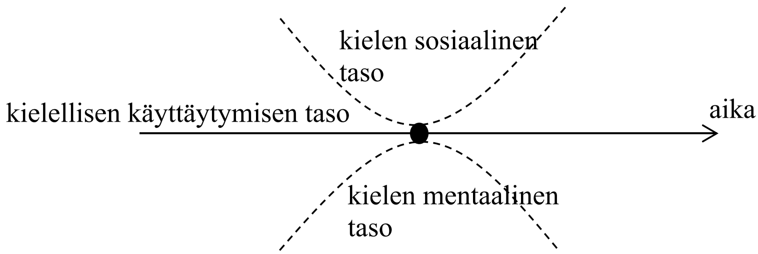
## 3.1. Kielikäsitykset lähtökohtana: esimerkkinä kieltä ei ole -teesi

Pysähdyttäessä havainnoimaan kieltä arkisessa elämässä huomataan, että valtaosa siitä koskee erilaisia ohimeneviä vuorovaikutus-, tarkoittamis- ja tulkintatilanteita, joissa varsinainen kielellinen toiminta (puhuminen, viittominen tai kirjoittaminen) on usein vain yksi osa viestinnän kokonaismodaliteettia (eleet, ilmeet, katseet, asennot jne.). Näkemys kielellisestä toiminnasta osana ohimenevää, yhä uudelleen vain kerran koettavaa tapahtumien virtaa on saanut amerikkalaisen uuspragmatistin Donald Davidsonin (2006 [1986], 265) esittämään kielifilosofiansa perustaksi niin sanotun kieltä ei ole -teesin (engl. *no language -thesis*), jonka mukaan yksittäistä kieltä ei ylipäätään ole siinä merkityksessä kuin erityisesti filosofiassa on totuttu ajattelemaan (nim. sisäsyntyisenä järjestelmänä). Näin Davidson toteaa:

> I conclude that there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed. There is therefore no such thing to be learned, mastered, or born with. We must give up the idea of a clearly defined shared structure which language-users acquire and then apply to cases. And we should try again to say how convention in any important sense is involved in language; or, as I think, we should give up the attempt to illuminate how we communicate by appeal to conventions. (Davidson 2006 [1986], 265.)

Davidsonin mielestä kielessä on konventioita, mutta ne eivät määritä tai määrää merkityksiä kielenkäyttötilanteissa. Tulkinta liittää ilmaisuihin merkityksiä, jotka viittaavat maailmaan ja käsillä olevaan vuorovaikutustilanteeseen, ja juuri tässä mielessä vuorovaikutuksen kulku ja onnistuminen tai epäonnistuminen ei ole ennalta saneltu. Tulee pyrkiä olemaan ymmärrettävä ja ymmärtämään muita. Viestinnällinen selviäminen edellyttää kielellisen toiminnan onnistumista riittävässä määrin. Ymmärrystä tai onnistumista ei kuitenkaan luo mikään ennalta asetettu sääntö tai konventio vaan alati uudistuvissa yksittäisissä

vuorovaikutustilanteissa tilapäiset, ohimenevät ja vastavuoroiset arvaukset puhekumppanien keskinäisistä intentioista. Tähän liittyy hänen käsitteensä radikaalista tulkinnasta, joka viime kädessä luo kommunikaation ja kielen merkitykset. Tätä jatkuvaa intentioiden arvailua Davidson kutsuu nimellä *passing theory*, ohimenevien tilapäisten oletusten teoria. (Ks. myös Seppänen 2004, 32–59.) Kuvio 15 pyrkii havainnollistamaan tätä tilannetta.




Kuvio 15. Kielen arkinen tilapäisten oletusten teoria ja kielen eri abstrakteja olemustyyppejä.

Ihmisten arkisessa elämänvirrassa kielellinen vuorovaikutus on kuvion 15 ilmentämää toimintaa ajassa. Musta pallo kuvaa yksilöä, jonka kielellinen toiminta ja käyttäytyminen on aikapaikkaista ja ehdoiltaan ja edellytyksiltään alati muuttuvaa ja uudistuvaa. Vuorovaikutuksen kannalta tavoitteena on tulla ymmärretyksi ja ymmärtää. Useimmiten kielellä on lähinnä tällainen yksilöiden välinen välinearvo tai välitön käyttöarvo tilanteisesti. Sillä pyritään selviämään arjen tapahtumissa. Tällaisen toiminnan voi hahmottaa kieltä ei ole -teesin mukaisesti.

Kuitenkin kielentutkimuksen perinteessä kieltä on tutkittu muutenkin kuin aikapaikkaisena toimintana ja käyttäytymisenä. Esimerkiksi kieli on voitu nähdä yksilön sisäistämänä mentaalisena järjestelmänä tai yksilöiden yllä vallitsevana sosiaalisena konventiona (kielen mentaalinen ja sosiaalinen taso kuviossa 15). Myös kielen eri olemusmuotojen keskinäisiä suhteita on voitu pohtia. Tällöin kielestä, sen olemuksesta ja näistä saatavasta tiedosta tehdään erilaisia oletuksia. Näiden oletusten kokonaisuus muodostaa kielen olemusta kuvaavan käsitejärjestelmän.

Kielentutkimuksen metateorian tehtävä on muun muassa tehdä selkoa tällaisista kielen teorioista.

Kieltä ei ole -teesi edustaa tässä yhtä esimerkinomaista näkemystä kielestä eli kielikäsitystä. Se osoittaa, että kieleen voidaan suhtautua hyvin monella eri tavalla, jopa kieltäen se sellaisena kuin kielentutkimus on kieltä perinteisesti hahmottanut.[1] Se on myös hyvä muistutus siitä, että kielentutkimuksella ei ole mitään suoraa ja välitöntä suhdetta tutkimuskohteeseensa, vaan erilaiset teoriat, suuntaukset ja metodiset toimintatavat tuovat esiin kielestä eri puolia tarkasteltavaksi. Ne sekä objektivoivat että kuvauksellaan luovat kielellisten ilmiöiden yleispiirteitä (ks. Määttä 2000a, 216–218; 2000b). Samalla niiden sisään rakentuu jokin teoreettinen näkemys kielellisestä todellisuudesta, mikä on tutkimuksessa hyvä tiedostaa. Kieltä ei ole -teesin virike kielentutkimukselle lienee siinä, että se pakottaa pohtimaan kohteen ja kuvauksen, empiirisyyden ja teoreettisuuden, kielellisen ilmiön ja kuvausterminologian jne. välistä suhdetta tutkimusta tehtäessä. Davidsonin näkemys, että *kielen* käsite voi johtaa tutkimusta harhaan on myös monimenetelmäistä kielentutkimusta ajatellen tärkeä aspekti, jota avataan seuraavaksi.

## 3.2. Ontologia ja kielentutkimus

Tutkimus ei ole koskaan teoriatonta. Metodin käyttö ja perustelu ovat aina sidoksissa oletuksiin todellisuuden luonteesta. Metodilla on siis käytännöllis-teknisen luonteensa lisäksi teoreettinen perustansa. Esimerkiksi 1970-luvulla keskusteltiin kiivaasti siitä, mikä on kielitieteessä kielijärjestelmän sisäsyntyisyyden tieto-opillinen asema eli toisin sanoen, mikä on generatiivisen transformaatiokieliopin luonne. 1980-luvun lopulla tällainen metateoreettinen keskustelu kehittyi ja väljentyi vastakkainasetteluksi funktionalistisen ja formalistisen koulukunnan välillä. Esimerkiksi Lakoff (1987) ja Chomsky (1989) esittivät kummatkin sen saman käsityksen, että eri koulukuntien näkemys kielen ontologiasta tai olemuksellisuudesta ja tutkimusmetodologia implikoivat toinen toisiaan. Myös esimerkiksi Langackerin (1987) mukaan hänen näkemyksensä kognition toiminnasta olisi suoraan sidoksissa

niihin kognitiivisiin tutkimuskäsitteisiin, joilla hänen käsityksensä mukaan kieltä tuli tutkia. Chomsky (1986) taas puolusti uusia formaalisia työkaluja omassa koulukunnassaan esittämällä saman argumentin.

Nämä keskustelut ja niissä esitetyt kysymykset vaikuttavat ehkä nykyisen suomalaisen metateoreettisen keskustelun näkökulmasta vanhentuneilta, sillä empiirisellä ja funktionaalisella kielentutkimuksella on ollut vahva perinne Suomessa. Maissa, joissa generativismi on edelleen vahvasti esillä kielentutkimuksessa, debatti on yhä ajankohtainen. Esimerkiksi Scholz, Pelletier ja Pullum (2016) esittelevät kielitieteen filosofiaa erityisesti generativismin näkökulmasta. He jakavat kielentutkimuksen harjoittajat kolmia: essentialisteihin, emergentisteihin ja eksternalisteihin. Näitä suuntauksia määrittävät asenteet tutkimuskohteeseen, tutkimuspäämääriin ja tutkimuksen tuloksiin. Phillipsin (2009) mukaan näiden koulukuntien välillä ei ole aitoa filosofista keskustelua: ”[d]ialogy between adherents of different approaches is alarmingly rare”.

Onpa kielitieteen erilaisista jaotteluista samaa tai eri mieltä, ne kuvaavat kieltä koskevien peruskäsitysten eli ontologian ja metodologian välistä kiinteää suhdetta. Ilman käsityksiä kielen olemuksesta ei voida pätevästi määritellä ja kuvata, miten tätä olemusta tutkitaan, saati sitä, mitä ja miten siitä pyritään tietämään. Oletukset tutkimuksen kohteesta johtavat erilaisiin tutkimusasetelmiin, metodeihin ja tulosten tulkintoihin. Kielitiede on monimenetelmäinen ala, ja jokaiselle tutkimusnäkökulmalle on tarkoituksenmukainen käyttö- ja sovellusalansa. Jokainen tässä teoksessa esitelty metodi tuottaa omasta näkökulmastaan rikasta ja kokonaiskuvaa täydentävää tietoa kielestä. Kielitiede voidaan siis nähdä Scholzin, Pelletierin ja Pullumin (2016) esitystä heterogeenisempana ja yhtenäisempänä. Tiukkoihin käytännön tutkimustyötä rajoittaviin rajanvetoihin ei ole välttämättä tarvetta, mutta erilaisten erontekojen heuristista arvoa tieteenalan kokonaishahmottamisen kannalta ei myöskään voi väheksyä.

Kielitieteen tieteenalan monimenetelmäisyyttä ajatellen on kuitenkin mielekästä ymmärtää kieli ilmiönä, jota voi käsitteellistää eri näkökulmista (ks. Arppe & Järvikivi 2007). Kun kieltä kuvataan eri näkökulmista ja lähtökohdista (esim. kuvion 15 kielen eri tasot), kohteen kuvauksetkin eroavat toisistaan. Kielitieteessä eri lähestymistavat eivät

siten ole lähtökohtaisesti toisensa poissulkevia, kuten saatetaan ajatella, vaan usein toisiaan täydentäviä. Näennäiset ristiriitaisuudet mallien, teorioiden ja kuvaustapojen välille syntyvät tavallisesti silloin, kun yleistetään yhdestä tutkimuksesta tai tutkimussuuntauksesta koko kieleen eli väitetään, että yksi näkökulma ilmiöön määrittää koko kielen. Tällöin tutkimusnäkökulman ja metodin aiheuttamat vaikutukset kohteen käsitteellistämiseen ovat hävinneet näköpiiristä. Kokonaisuutta tavoitteleva systeemiteoreettinen näkökulma monimenetelmäisyyteen on helppo ymmärtää elefantti-faabelin kautta: siinä kolme sokeaa miestä päätyy ilmiriitaan siitä, mikä on elefantti. Kun yksi on tarttunut kiinni hännästä, toinen jalasta ja kolmas kärsästä, tuntuu, että kuvaillaan eri asioita. Kielitieteessä samoin kielitajun varainen autonominen lingvistiikka, keskustelutallenteiden varainen vuorovaikutuslingvistiikka tai laajoihin tekstiaineistoihin perustuva korpustutkimus eivät poissulje toisiaan vaan lisäävät kukin, tarkoituksenmukaisesti sovellettuina, omasta näkökulmastaan tietoa kielestä.

Tutkimuksen näkökulman tiedostamisessa on oleellista ymmärtää oman tutkimuskäsitteistön sidoksisuus niihin ontologisiin oletuksiin, jotka tutkimuksen aikana tehdään. Tämä ei ole kielitieteelle erityinen ominaisuus vaan pätee yleisesti tieteelliseen tutkimukseen. Kielitiede ei siten tältä osin eroa muista tieteistä, vaan eri tieteenalat ovat tässä suhteessa samankaltaisia. Kuten aiemmin esitimme (ks. Johdannon luku 2.5. tk.), tieteellinen tutkimus on aina teoriapitoista; ”puhdasta” havaintoa ei nykykäsityksen mukaan saavuteta, vaan tutkimustermien merkityksen määrää käytössä oleva taustateoria. Quinen (1951) holistisen merkitysteorian klassikko esittää, ettei riippumatonta väitelausetta voi verrata todellisuuteen, vaan taustalla on aina jokin koherentti uskomusjärjestelmä. Niinkin yksinkertainen ja perustava kielitieteellinen käsite kuin *merkki* saa erilaisen merkityksen de Saussuren, Peircen, Wittgensteinin ja Chomskyn kirjoituksissa. Jos taustalla olevat uskomusjärjestelmät ovat erilaisia, väitelause ”kieli on merkkien systeemi” tarkoittaa eri asioita. Kielentutkijan on siten oleellista pohtia, missä merkityksessä hän käyttää omia peruskäsitteitään, millaisia sidoksisuuksia erilaisiin uskomusjärjestelmiin niillä on ja mitä loogisia implikaatioita eri merkityksillä näiden sidoksisuuksien myötä on. Hyvä on myös

pohtia sitä, millä tavalla oma määritelmä kohteesta mahdollisesti rajaa tai sulkee pois jonkin toisen alan näkökulmia kielentutkimukseen.

Tutkimuksen lähtökohtaisella teoriasidonnaisuudella on erilaisia seurauksia. Reaalitieteissä, joihin kielentutkimuskin pääosin kuuluu (ks. Johdannon luku 2.3. tk.), on tutkimuksen kuluessa otettava eksplisiittisesti tai implisiittisesti kantaa siihen, millaiseen todellisuuden alueeseen tutkimus kohdistuu (ks. kuviossa 15 kielen eri olemustyypit). Tutkimuksen kohde hahmotetaan osana laajempaa todellisuutta. Tutkimusasetelmaa ei voi suunnitella, ellei ole tällaista esiymmärrystä ja ennakkokäsitystä siitä, mikä tutkimuskohde yleispiirteiltään on, miten se liittyy maailmassa muihin ilmiöihin ja miten siitä voi saada tieteellistä tietoa.

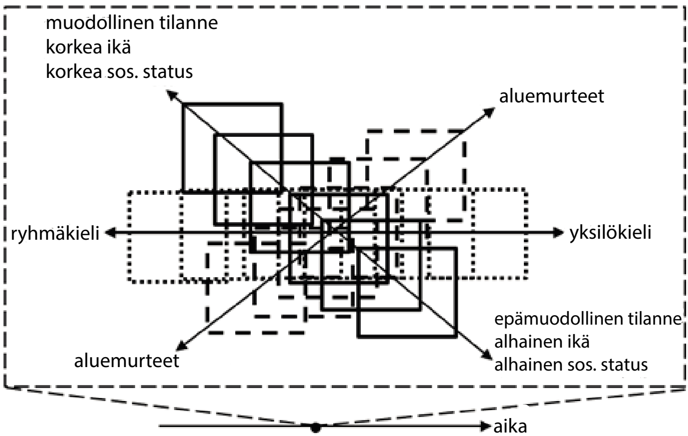
Kielitieteessä ontologialla on erityisen merkittävä rooli: kielentutkijan on muodostettava käsitys kahdesta eri kielestä, joista toinen on tutkimuskohde, luonnollinen kieli eli objektikieli, ja toinen tutkimuskohdetta koskeva kielellinen esitys, tieteen kieli eli metakieli. Kun tutkijan on monilla muilla aloilla ainoastaan vastattava kysymykseen, mikä on tieteen kielen ja todellisuuden suhde, kielitieteilijän on vastattava myös kysymykseen, onko tieteen (meta)kielellä ja luonnollisella, tutkittavalla objektikielellä sama ontologia ja mikä ontologia näillä kahdella mahdollisesti on. Jos kielentutkija esimerkiksi esittää, että tieteen kieli ei viittaa todellisuuteen, hylkää hän myös samalla usein tieteellisen realismin (ks. Johdannon luku 2.2. tk.) ja totuuden korrespondenssiteoriat (ks. Johdannon luku 2.4. tk.), eivätkä nämä tieteenfilosofiset taustateoriat voi tällöin olla osa tutkimuksen teoreettista viitekehystä. Johdannon luvussa 3.2.2. tarkastellaan pintapuolisesti käsitteitä *laki* ja *selitys* erityisesti suhteessa realistiseen todellisuuskäsitykseen, jotta lukija saisi paremman kuvan siitä, millaista epäkoherenssia tai millaisia loogisia ristiriitaisuuksia aivan tavalliset kielitieteen termit voivat tutkimuksen kuluessa aiheuttaa.

### 3.2.1. KIELEN OLEMUS EI OLE SAMA KUIN KIELEN ONTOLOGIA

Tutkimus katsoo kohdettaan ja käsitteellistää sitä aina tietystä näkökulmasta. Eri näkökulmat, metodit ja aineistot havaitsevat siten ilmiöistä eri puolia. Juuri yksittäisten tutkimusten näkökulmasidonnaisuus tekee monimenetelmäisestä kielentutkimuksesta arvokkaan. Siksi kaikkia menetelmiä tarvitaan (ks. Arppe & Järvikivi 2007a). Monen

tutkimusnäkökulman rakentaminen yhteen kohteeseen ei kuitenkaan implikoi, että kohde olisi ontologisesti moninainen, vaan että tutkimus on rikasta. Toisin sanoen erilaiset tutkimuskohteen tasot, rakenteet ja suhteet ilmenevät erilaisille tutkimusnäkökulmille usein erilaisina, ja tutkimuksissa käytetään tällöin usein myös eri käsitteitä.

Ajatus, että kieli on systeeminen kokonaisuus, on yleensä tuttu kielitieteilijälle. Termi, sääntö, konventio, muuttuja tms. voidaan viime kädessä kuvata vain toisten termien, sääntöjen, konventioiden ja muuttujien avulla, jotka siis yhdessä muodostavat jonkinlaisen kokonaisjärjestelmän. Kuvio 16 havainnollistaa puhutun ja viitotun kielen olemuksellista moninaisuutta kolmen muuttujadimension avulla (ks. myös Karlsson 1982, 49). Muuttujat ovat tyypillisiä esimerkiksi perinteiselle (labovilaiselle) sosiolingvistiselle tutkimukselle, ja ne ilmentävät esimerkinomaisesti kieltä systeeminä, jossa tutkitut muuttujat tai tekijät muodostavat rajatun selitysavaruuden jostakin kielen ilmiöstä.




Kuvio 16. Puhutun ja viitotun kielen vaihteluun vaikuttavia muuttujia (mukaillen Karlssonia [1982]).

Kieltä voi tieteessä tarkastella esimerkiksi aluemurteiden, perinteisten sosiolingvististen muuttujien (sukupuoli, ikä, sosiaalinen luokka) ja ryhmäidentiteetin kannalta. Kukin muuttujaulottuvuus luo valitun

näkökulman ja usein myös käytettyjen metodien myötä hahmottaa oman, jollain tavoin rajoittuvan kielimuotonsa: esimerkiksi Turun seudun murre, kulosaarelaisen korkeakoulutetun kieli, Liisin (synt. 1938 Leppävirta) kieli tai katutanssiryhmän kieli. Tällainen hahmotus luo mitä moninaisimpina risteävien ja limittyvien kielen muotojen tai systeemien verkoston, jossa mikään kohta tai leikkaus yksinään ei edusta kieltä sinänsä ja holistisesti. Kun kielen ilmentymien (esim. kirjoitettu ja puhuttu kanava) ja muuttujien limittymisen lisäksi ilmaustapojen valinta ajassa ja kielellinen toiminta otetaan mukaan yhtenä vaikuttavana tekijänä, osoittaa se, että on hyvin vaikea hahmottaa mitään itsenäistä ja autonomista yksikköä kielessä tai rajautuvaa kielioliota, esimerkiksi suomen kieltä (ks. Johdannon luku 3.1. tk.).

Toisinaan kielentutkimuksen metodologiaa käsittelevässä kirjallisuudessa esiintyy ontologisen pluralismin yhteydessä Popperin (1978) kolmen maailman malli. Popperin mallissa lyhenteet M1, M2 ja M3 viittaavat todellisuuden kolmeen erilliseen maailmaan: fysikaaliset oliot ja tapahtumiset (M1), mentaaliset oliot ja tapahtumiset (M2) sekä ihmisten ja yhteisöjen ajattelulla tuotetut sisällöt (M3, esim. tieteelliset teoriat, sosiaaliset instituutiot, taideteokset) (ks. esim. Itkonen & Pajunen 2010, 11–12). Mallia tarkasteltaessa on mielekästä kuitenkin kiinnittää huomiota siihen, että malli toimii Popperin **kriittisen rationalismin** perustana.[2]

Popperin mallia käytetään heuristisesti myös ontologialtaan monistisissakin teorioissa. Esimerkiksi **fenomenologiset teoriat** esittävät, että M1, M2 ja M3 ovat olemuksellisen ilmenemisen tasoja niin ihmiselle kuin myös tieteessä. Myös foneettiset mittaukset tai kielellistä aivotoimintaa kuvantavat menetelmät edustavat erilaisia tarttumapintoja kieleen, ja niitä voi kuvata Popperin M1:n kautta tai jonkin muun kilpailevan käsitteistön avulla. Teorialle, jossa on fenomenologinen epistemologinen näkökanta, kielen keskenään erilaisten olemustyyppien ja ilmiöiden tutkiminen ei siten tarkoita erillisten ontologisten tasojen tutkimusta.

Monen näkökulman rakentaminen tutkimuskohteeseen ei siten suoraan implikoi, että kohde olisi ontologisesti pluralistinen vaan että ihmisen rakentama kokemusmaailma siitä on rikas. Toisin sanoen erilaiset maailmat, tasot, rakenteet, elämänmuodot ja suhteet ilmenevät ihmiselle

ja erilaisille tutkimusnäkökulmille tutkimuksen eksplikoinnin myötä. Ontologinen monisti, niin fenomenologi kuin materialistikin, voi siten kannattaa tutkimusmetodien moninaisuutta tieteellisen tiedon kartuttamisen välineenä. Tällaisen näkemyksen Popperin kolmeen maailmaan esittävät kielentutkimuksessa esimerkiksi Lakoffin ja Johnsonin (1999) ruumiillinen realismi ja prototyyppiteorian taustalla oleva neurofenomenologia (Varela, Kabat-Zinn, Rosch & Thompson 2017).

Suomalaisessa tieteenfilosofiassa (esim. Niiniluoto 1990) ja kielitieteessä (Itkonen 1997, Määttä 2000c) vallalla ollut ontologinen näkemys on vanhastaan ollut pluralistinen. Nykytutkijan on kuitenkin hyvä tiedostaa, että nykyisin juuri pluralismi tieteenfilosofisena näkemyksenä on haasteiden edessä, ja sitä näyttää olevan erittäin vaikea puolustaa. Johdannon luvussa 2 on käsitelty tähän metafyysisen filosofian kehitykseen vaikuttaneita seikkoja, kuten vaikeuksia selittää substanssien väliseen interaktioon liittyvää kausaalista tai muuta yhteyttä. Ontologiset pluralistit (esim. Putnam, Hirsch) usein esittävät, että on olemassa eri kieliä, joiden avulla eri olemassaolon tyyppejä voidaan esittää, ja että nämä kielet voivat kuvata maailman tosiasioita periaatteessa yhtä hyvin. Ontologiset kysymykset olisivat siis pääasiassa kielellisiä, ja ne palautuisivat kielellisiin valintoihin. Miten tällainen näkemys soveltuu kielitieteeseen itseensä ja mitä implikaatioita sillä olisi kielentutkimuksen metodologiassa? Tervetullutta olisikin uusi 2000-luvun kielitieteen filosofia, joka ottaisi huomioon pluralismin sudenkuopat ja esittäisi uudenlaista koherenttia kuvaa kielestä ja sen monimenetelmäisestä tutkimuksesta.

### 3.2.2. TODELLISUUSKÄSITYKSEN VAIKUTUS TUTKIMUKSEEN: LAKI JA SELITTÄMINEN KIELENTUTKIMUKSESSA

Vastaus esimerkiksi kysymykseen ”Onko kielen normi samaa vai eri substanssia kuin materiaalinen maailma?” määrittää kielentutkijan positiota suhteessa siihen, koskevatko kieltä erityiset omat tutkimusmetodit. Teoriat voidaan lisäksi jakaa sen suhteen, miten ne vastaavat kysymykseen, mikä on todellisuuden suhde tietoisuuteen: Onko kielen normi olemassa silloinkin, kun yksikään ihminen ei sitä ajattele, vai onko se riippuvainen tietoisuudesta? Löydetäänkö kielen piirteet todellisuudesta vai konstruoidaanko ne? Vastaukset määrittävät käsityksiä siitä, miten

kieltä tulisi tutkia. Tässä luvussa käsitellään käsitteitä *laki* ja *selittäminen* kielentutkimuksessa suhteessa todellisuuskäsityksiin.

Käsitteellä laki voidaan viitata 1) luonnollisen maailman fysikaalisten ilmiöiden kausaaliseen selittämiseen eli seuraantolakeihin, 2) eksperimentaalisiin lakeihin (esim. veden kiehumispiste), 3) empiirisiin lakeihin (esim. Galilein putoamislaki), 4) teleologisiin selityksiin (A teki B:n, koska tavoitteli C:tä), 5) ilmiöiden invariantteihin muuttumistapoihin (esim. Darwinin luonnonvalinta), 6) funktionaalisiin riippuvuussuhteisiin (mentaalisia toimintoja ei ole ilman muutoksia fysikaalisessa todellisuudessa), 7) kvalitatiivisiin lakeihin (korpit ovat mustia), 8) yhteiskunnan velvoittaviin säädöksiin ja juridiseen säätelyjärjestelmään (esim. Suomen rikoslaki), 9) teoreettisiin lakeihin ja niitä kuvaaviin lauseisiin (”Samalla aineella samassa tilassa on samat ominaisuudet”) ja 10) universaaleihin piirteisiin (esim. kaikissa kielissä on vokaaleja). Termillä *laki* voidaan siis viitata hyvinkin erilaisiin ilmiöihin. Yleensä erilaiset tieteenfilosofiset teoriat pyrkivätkin selittämään, miten nämä erilaiset laki-käsitteeseen kuuluvat ilmiöt liittyvät toisiinsa. Ne nimeävät käsitteet selvyyden vuoksi usein eri termein.

Ottaen huomioon, että käsite laki ja siihen liittyvä termistö ovat kiistatta tieteenfilosofian hankalimpia käsitteitä, joita on jo Aristoteleesta asti käsitelty, on varsin valitettavaa, että vielä 2000-luvulla ei ole perinpohjaista kielitieteen filosofiaa, jossa olisi käsitelty erilaisten laki-käsitteiden käyttöä kielentutkimuksessa. Keskustelu lainomaisuuksien ja säännönmukaisuuksien luonteesta kielentutkimuksessa johtaa usein sellaiseen virhepäätelmään, että *laki*-termillä olisi vain yksi merkitys (esim. ’deterministinen luonnonlaki’) tai että vain luonnontieteissä olisi olemassa empiirisiä lakeja ja ihmistieteissä pelkästään hermeneuttista tulkintaa. (Ks. Tieteen termipankki s.v. laki.)

Tiedostamattomat näkemykset todellisuuden luonteesta nivoutuvat usein tutkijan huomaamatta tutkimuskäsitteistöön ja ohjaavat uskomuksia ja päätelmiä tutkimuksen kohteesta. Käyttämällä yhtä tai toista termiä hän voikin tietämättään vastata seuraavanlaisiin kysymyksiin: Mikä on tieteellisen selityksen rakenne? Onko selittäminen ja ennustaminen sama asia? Onko luonnontieteen selitys sama kuin ihmistieteen selitys? Miten selitysvoimaa voi arvioida? Mitkä ovat lain lainomaisuuden

tunnusmerkit? Mitä eroa on tosiasioiden selittämisellä ja lakien selittämisellä? Onko olemassa todennäköisyyksiä koskevia lakeja?

Lait voidaan karkeasti jakaa a) seuraantolakeihin ja b) yhteenliittymislakeihin. Deterministiset lait, kuten ne, jotka Newtonin mekaniikka ilmaisee, ovat seuraantolakeja. Laki ilmaisee syyn ja seurauksen, jotka toteutuvat aina standardiolosuhteissa. Tilastolliset yhteenliittymislait, esimerkiksi perinnöllisyyslait, taas kertovat, millä todennäköisyydellä jokin ilmiö tapahtuu (esim. lapsi on 51 %:n todennäköisyydellä poika). Seuraantolait kuvaavat siis tapahtumien ja niiden kulun säännönmukaisuutta, yhteenliittymislait joidenkin piirteiden, ominaisuuksien tai suhteiden säännönmukaista yhteenliittymistä (ks. Niiniluoto 1983, 229–236).

Termeillä *lainomainen* ja *satunnainen* erotellaan sellaiset lauseet, joilla on jokin laki-käsitteen määritelmään kuuluva piirre, ja ne lauseet, jotka ovat satunnaisia. Termillä *lainomainen* viitataan sellaisiin lauseisiin, jotka johtavat suoraan tai välillisesti informatiivisiin yleistyksiin, kuten ”huoneen dimensiot ilmaisevat, kuinka paljon happea voi tässä tilavuudessa olla”. Myös esimerkiksi luonteeltaan tilastolliset seuraantolait (esim. todennäköisyys saada rahanheitossa kolme kruunaa peräkkäin on 1/8) tai universaalit yhteenliittymislait (aggressiiviset henkilöt ovat väkivaltaisia) ovat lainomaisia. Sen sijaan esimerkiksi lause ”kaikki koululuokassa ovat syntyneet vuoden 1990 jälkeen” voi ilmaista toden yleistyksen, mutta se ei ole lainomainen vaan satunnainen eli kontingentti (ks. Johdannon luku 2.3. tk.). Tieto, että huoneessa on tiettyyn ikäryhmään kuuluvia ihmisiä, ei johda sellaisiin yleistyksiin, jotka lisäävät tietoa huoneista tai ihmisryhmistä yleensä.

Termillä *lainomainen* voidaan myös viitata siihen, että taustateoriassa ei uskota, että deterministisiä lakeja on ollenkaan olemassa, eli *laki*-termillä voidaan nimetä oikeastaan ainoastaan lainkaltaisia ilmiöitä (ns. *ceteris paribus* -käsitteestä (ks. Johdannon luku 2.5. tk.). Humelainen käsitys laeista on, että on olemassa vain tapahtumien tosiasiallisia yhteenliittymiä. Ilmaus ”B seuraa A:ta” ilmaisee vain ja ainoastaan sen, että jotkin asiat ovat historiallisesti liittyneet yhteen. Popperinkaan (2002 [1934]) mielestä aitoja seuraanto- tai kehityslakeja ei ole olemassa. Esimerkiksi Niiniluodon (1990) mukaan lainomaisilla lauseilla on asteita. Voidaan siis kysyä, missä määrin jokin ilmiö on lainomainen.

Taulukossa 1 on käytetty Niiniluodon (1983, 204) jakoa seuraanto- ja yhteenliittymislakeihin, jotka vuorostaan voidaan jakaa teoreettisiin ja empiirisiin, deterministisiin ja tilastollisiin sekä kvalitatiivisiin ja kvantitatiivisiin lakeihin. Empiirisiksi laeiksi kutsutaan sellaisia lakeja, jotka voidaan todentaa havainto- ja kokemusaineistolla (esim. Galilein putoamislaki). Eksperimentaaliset lait ovat empiiristen lakien eräs tapaus, sillä niissä on kokeellisesti mitattavissa olevia kvantitatiivisia käsitteitä (esim. veden kiehumispiste). Teoreettisilla laeilla viitataan ei-havaittaviin ominaisuuksiin. Deterministiset lait ovat usein dynaamisia välttämättömiä lakeja, ja tilastolliset lait vuorostaan kertovat, millä todennäköisyydellä jokin asia tapahtuu. Niiniluodon (mp.) kaavio toimii heuristisena välineenä sen ymmärtämiseksi, että käsitteiden *laki* ja *lainomaisuus* alla pohditaan useita loogisilta ominaisuuksiltaan erilaisia ilmiöitä.

| | | C) deterministinen | C) deterministinen | E) tilastollinen | E) tilastollinen |
|---|---|---|---|---|---|
| | | kvantitatiivinen | kvalitatiivinen | kvantitatiivinen | kvalitatiivinen |
| A) seuraantolaki | teoreettinen | | | | |
| A) seuraantolaki | empiirinen | dynaaminen laki | | | |
| B) yhteenliittymislaki | teoreettinen | | Kaikissa luonnollisissa kielissä on vokaaleja ja konsonantteja. | | Jos kielessä neutraali lausekejärjestys on SOV, siinä on ensisijaisesti postpositioita. |
| B) yhteenliittymislaki | empiirinen | | Kaikki korpit ovat mustia. | Zipfin laki | äännelaki |

Taulukko 1. Lait ja lainomaisuudet Niiniluotoa (1983, 204) mukaillen.

Kuten huomataan, lainomaiset ilmiöt voidaan karkeasti jakaa 16 luokkaan. Näistä vain kaksi sisältää klassiset universaalit seuraantolait ja yksi dynaamiset lait, jotka useimmiten nimetään *luonnonlaeiksi*. Dynaamiset lait ilmaistaan ajallisena funktiona (esimerkiksi Galilein putoamislaki ½ $gt^2$), ja ne muodostavat siten laki-kategoriassa erittäin pienen joukon. Vaikka taulukko 1 ei ole kannanotto Niiniluodon (1999) kriittisen tieteellisen realismin puolesta, tämä yksikertainen käsitteistö toimii hyvänä perustana keskustelulle kielitieteessä käytettyjen erilaisten yleistysten, säännönmukaisuuksien ja lakikäsitteistön ominaisuuksista. Taulukkoon 1 on sijoitettu muutamia erilaisia kielitieteellisiä laki-käsitteitä siinä toivossa, että jako herättäisi vilkkaampaa keskustelua kielitieteen omasta lainomaisuuskäsitteistöstä. Taulukkoa 1 voi käyttää myös oman arkitutkimuksen käsiteanalyysissa. Käsitteen käytön ala nimittäin kertoo myös sen, millaisia piirteitä ilmiöllä oletetaan olevan ja mikä siten on tyydyttävä ilmiötä koskeva tieteellinen selitys.

Tieteelliset selitykset ovat yksi tieteellisen argumentaation muoto. Edellä luvussa 2 käsiteltiin tieteellisen argumentin rakennetta ja todettiin, että argumentti koostuu väitteestä ja perusteluista. Tieteelliset argumentit voidaan jakaa karkeasti kahteen perusluokkaan reaalitieteissä: 1) kuvauksiin, jotka kertovat, mitä maailmassa on, ja 2) selityksiin, jotka kertovat, miksi jokin asia on niin kuin se on. Kummassakin tapauksessa tieteellisen argumentin vaikuttavuus liittyy väitteen perusteluihin ja näyttöön.

Kielitieteessä kuvaamisen tavoite on kertoa, millainen kieli on, ja selityksellä tavoitellaan sitä, miksi kieli on, mitä se on. Kuvauksessa käytetään luokittavia metatermejä. Tavoite on yhtäältä a) kuvata reaalitodellisuuden kieli-ilmiöitä ja toisaalta b) pyrkiä kehittämään ja tarkentamaan ilmiöitä nimeävien metatermien määritelmiä erilaisten aineistojen, esimerkkien ja uuden tieteellisen tiedon valossa.

**Kielitieteellinen kuvaus**

a) Kuvauksessa tavoitteena on kuvata reaalitodellisuuden kieli-ilmiöitä niin aivojen, kognitiivisten toimintojen, puheen, käytettävien merk-

kien, kielellisen toiminnan, kieliopin järjestelmien kuin kirjoitetun kielen tasolla.

b) Kielenkuvauksen tavoitteena on kehittää edelleen metateoreettisia luokitusjärjestelmiä niin, että ne kattavat tunnetut aineistot ja ilmiöt ja että ne luokittavat ilmiöitä mahdollisimman hyvin.

Kielitieteellisten selitysten voidaan katsoa muodostavan yhden keskeisimmistä tieteenalan tehtävistä: pelkkä onttinen luettelo (mm. kielessä on verbejä, substantiiveja jne.) ei ole tiedollisesti tyydyttävä. Kiinnostavaa on tietää, miten erilaiset kielen sisäiset ja ulkoiset ilmiöt liittyvät toisiinsa ja millaiset ilmiöt aiheuttavat tapahtumaketjuja kielessä ja kieliyhteisöissä. Tieteellisen vallankumouksen voidaan katsoa liittyvän juuri siihen, että ymmärrettiin, että selittävä tutkimus mahdollistaa ymmärtämisen ja tulevien tapahtumien ennustamisen, kontrollin ja erilaiset teknologiat.

Toisinaan tieteellinen selitys assosioidaan yksinomaan laki-käsitteeseen: ajatellaan, että tieteellisissä selityksissä on kysymys vain jonkin lain (yleensä dynaamisen luonnonlain) kuvaamisesta tai deduktiivis-nomologisesta selittämisestä (ks. subsumptioteoria ja DN-malli Johdannon luku 2.5. tk.). Ajatus, että jonkin tapahtumainkulun taustalla on invariantti (muuttumaton) universaali laki, on kuitenkin verraten uusi. Aristoteles, joka aloitti keskustelun tieteellisestä selittämisestä, esitti, että selitykset koskevat ilmiöiden **substantiaalista invarianssia** eli olemuksellista pysyvyyttä tai muuttumattomuutta. Aristoteelinen maailmankäsitys oli vallitseva aina valistusajan tieteelliseen vallankumoukseen eli 1600-luvulle asti. Aristoteelinen selittäminen on luonteeltaan teleologista ja perustuu ajatukseen, että ilmiöillä on jokin essenssi, niille ominainen luonne, ja siten tieteelliset selitykset kertovat erilaisten maailmaan kuuluvien olioiden pyrkimyksestä tavoittaa päämääränsä, potentiansa (Kaila 1939; von Wright 1971). Tieteellinen selitys kuvaa ilmiön ominaisuuksien ja potentian syitä.

Huomataan, että tieteellisellä selittämisellä on kiinteä suhde vallitsevaan maailmankäsitykseen. Käsitys tieteellisen selittämisen luonteesta

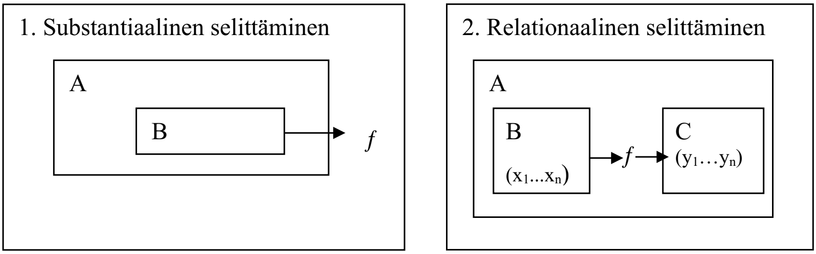
onkin kehittynyt rinnakkain tieteen kanssa. Kun käsitys maailman luonteesta muuttuu, muuttuvat myös käsitykset siitä, miten maailmassa olevia suhteita selitetään. Taulukosta 1 huomataan, että käsite *laki* sisältää ainakin 16 kategoriaa. Nykyinen käsitys laeista on niin Aristoteleen kuin Galileinkin laki-käsityksiä laajempi. Nykyään lainomaisiksi ilmiöiksi ymmärretään myös erilaiset teoreettiset kvalitatiiviset ja kvantitatiiviset yhteenliittymislait ja -suhteet.

Kielitieteen selityksillä on ollut joitakin alalle ominaisia piirteitä. De Saussuren strukturalismi on fonologian suhteen staattinen ja synkroninen mutta morfologian suhteen dynaaminen ja diakroninen. Kummassakin tapauksessa ilmiö selitetään suhteessa siihen järjestelmään, jonka osa se on. Autonomisessa kielitieteessä kielen formaalia kuvausta on myös pidetty luonteeltaan selittävänä. Tavoitteena on ollut paljastaa luonnontieteellinen noominen ilmiö eli biologinen universaalikielioppi, joka selittäisi kielen rakenteita. Generatiivinen formaali kuvaus on siis luonteeltaan samalla selitys, koska se kertoo, millainen universaali invarianssi on johtanut johonkin kieli-ilmiöön. Hyvin samankaltainen selittävä suhde on kognitiivisella kielitieteellä suhteessa kognitioon. Kognitio selittää kielen rakenteita eli vastaa kysymykseen, miksi kieli on sellainen, kuin se kuvausten perusteella näyttää olevan.

Kielitieteessä myös funktionaalisella selittämisellä on ollut tärkeä rooli. Funktionaalisessa selittämisessä jotain kielen ilmiötä koskevaan miksi-kysymykseen vastataan viittaamalla sen tehtävään eli funktioon (*f*) jossain laajemmassa kokonaisuudessa. Miksi sydän sykkii, tai miksi suomen kielen *y*-äänne esiintyy pyöreänä? Sydän sykkii, koska se pumppaa verta biologisessa systeemissä. Suomen *y*-äänne on pyöreä, jotta se erottuisi saman ääntymäpaikan laveasta *i*-äänteestä (Määttä 1994, 72).

Funktionaalista selittämistä on toisinaan pyritty ymmärtämään suhteessa aristoteeliseen teleologiseen selittämismalliin. Kielen rakenteen finaalisuus on tällöin yleistetty aristoteelisesta praktisesta syllogismista tiedostamattomaan ja kollektiiviseen toimintaan (Määttä 1994, 258). Tähän tulee suhtautua varauksella sikäli, että Aristoteleen käsitteistö ei kata **relationaalista invarianssia** eli suhteita koskevaa säännönmukaisuutta. Substantiaalisen ja relationaalisen sekoittaminen toisiinsa johtaa kielen mystifioimiseen: ”kielellä on päämääriä”, ”kieli reagoi

rakenteellisesti", "morfeemilla on muutospyrkimys" jne. Substantiaalinen selittäminen perustuu ilmiön ominaisuuksien (essenssin tai potentian) ja ulkoisen ympäristön suhteeseen. Tämä tarkoittaa, että ilmiössä B tapahtuvat muutokset on selitettävä toisaalta niillä tehtävillä, jotka sillä on ilmiössä A, mutta erityisesti niillä finaalisyillä tai päämäärillä, joita A:lla oletetaan olevan. Substantiaalinen selittäminen kuvaa B:n tehtävää suhteessa A:n finaalisyihin. (Ks. kuvio 17.)




Kuvio 17. Substantiaalisen ja relationaalisen invarianssin funktionaalinen selittäminen.

Relationaalisen invarianssin selittäminen vuorostaan on nimensä mukaisesti relaatio: funktio tässä mallissa on sääntö, joka liittää kaksi joukkoa yhteen siten, että $f$: B → C. Funktio liittää siis yksi yhteen määrittelyjoukon B:n alkiot ($x_1...x_n$) maalijoukon C:n alkioiden kanssa ($y_1...y_n$). (Ks. kuvio 17.) Relationaalisessa funktiokäsityksessä kummallakaan joukolla, B:llä tai C:llä, ei ole "persooniin assosioitua" intentiota, vaan puhutaan pelkästään funktion määrittelyjoukon argumenteista ja maalijoukon arvoista.

Relationaalinen funktiokäsitys on implisiittisesti käytössä tilastotieteessä, mikä tarkoittaa sitä, että niillä kielentutkimuksen alueilla, joissa tilastotieteillä on rooli, kuten sosiolingvistiikassa, psykolingvistiikassa ja typologiassa, tämä käsite on osa metodologiaa. *Relationaalisen funktion* käsite on käytössä myös kielitieteen ulkopuolella, esimerkiksi biologiassa, psykologiassa, erilaisissa itseohjautuvien systeemien teorioissa, kybernetiikassa, kaaosteoriassa ja kompleksisten ilmiöiden matematiikassa. Relationaalinen funktiokäsitys on aina oleellinen silloin, kun kompleksisuus ymmärretään systeemin luonnollisena ominaisuutena

ja kehitysehtona. Kompleksien systeemien määritelmä perustuu näissä teorioissa siihen, että osat ovat yhteydessä toisiinsa, kokonaisuuteen ja ympäristöönsä hyvin moninaisin eri funktioin. Tämä kytkeytyneisyys johtaa siihen, että osat vaikuttavat sekä toisiinsa että takaisin systeemikokonaisuuteen.

Kun substantiaalinen selitysmalli kuvaa ilmiön essentiaalisia ominaisuuksia muutoksen syinä, relationaalinen selitysmalli vuorostaan tarkastelee ilmiön kytköksien vaikutusten seurauksia eri näkökulmista. Kyseessä on siis kaksi hyvin erilaista selittämisen mallia. Toinen perustuu essenssiin eli olemukseen ja toinen relaatioihin, toimintoihin ja rakenteisiin.

**Tieteellisiä selitystyyppejä**

- noomiset selitykset (esim. luonnonlaeilla selittäminen)
- teleologiset selitykset (esim. ihmisen käyttäytymisen selittäminen tavoitteiden ja intentioiden valossa)
- historialliset selitykset (esim. ilmiön synnyn ja kehittymisen kautta selittäminen, evolutionaarinen ja geneettinen selittäminen)
- psykologiset selitykset (esim. selittäminen ihmisen ajattelun ja kognitiivisten toimintojen valossa)
- funktionaaliset selitykset (esim. tehtävän mukaan selittäminen)
- hermeneuttiset selitykset (esim. ilmiön ymmärtämiseen pyrkivä selittäminen)

Kuten edellä esitimme, nykykäsityksen mukaan havaitseminen on käsitteellistävää ja teoriapitoista (ks. Johdannon luku 2.5. tk.). Teoriapitoisuus on eri kysymys kuin kysymys ilmiöiden itsenäisestä olemassaolosta. Realistit olettavat, että todellisuus on havaitsijasta riippumatonta eli että puu on puu silloinkin, kun sitä ei katsele tai muulla tavoin havaitse kukaan tai mikään. Jos kielentutkija kokee itsensä realistiksi, on hänen hyvä pohtia, missä määrin kieli on osa ihmisestä riippumatonta todellisuutta eli missä määrin se on itsenäisesti olemassa. Vastakkainen näkökulma on esimerkiksi, että ilman tiedostavaa ja tulkitsevaa ihmismieltä

fysikaalisessa todellisuudessa on pelkästään merkityksettömiä visuaalisia kuvioita (vrt. kirjoitusmerkit) tai ääniaaltoja (vrt. puhe). Esimerkiksi konseptualismin mukaan kieli on käsitteellisesti riippuvainen merkeille tai ääniaalloille annetusta tulkinnasta eli ihmisestä ja tämän tajunnasta.

Tieteellisen realismin mukaan tiede kuvaa puuta paremmin kuin pelkkä ihmisen aistitieto. Niiniluodon (1993) mukaan tieteellinen kieli ei kehity tyhjiössä, vaan tieteissä kielen ja reaalitodellisuuden välille muodostuu toiminnallinen suhde, jossa tutkija muokkaa tieteen kieltä jollain tavoin reaalitodellisuutta vastaavaksi. Koska tieteellisiä väitteitä testataan ja koetellaan, ne eivät ole samanarvoisia kuin erilaiset arkipuheessa esitetyt väitteet ja uskomukset. Esimerkiksi kielitieteessä vertaileva tutkimusasetelma voi joko tukea lausetta ”*n*:nneksi käytetyimmän sanan frekvenssi missä tahansa luonnollisessa kielessä (jonkin kielen jossakin laajassa tekstissä) on käänteisesti verrannollinen lukuun *n*” tai kieltää sen. Tieteessä väitteiden totuusarvoa tutkitaan, ja siksi jotkin väitteet kelpaavat tieteelliseksi tiedoksi.

Edellä esitetty väite sanan käyttötiheydestä on George Zipfin (1935) havainto, jota kutsutaan usein *Zipfin laiksi*, mutta täsmällisempi termi on kuitenkin *Zipfin jakauma* (ks. Tieteen termipankki s.v. Zipfin laki). Zipfin jakaumaa esiintyy niin kielessä, populaatioissa, tulotasoissa kuin tv-kanavien katsojaluvuissa jne. Syitä siihen, miksi Zipfin jakaumaa esiintyy, on yhtä monta kuin ilmiöitä, joissa jakaumaa esiintyy. Zipf itse esitti, että kielessä jakauma muodostuu, koska ihmiset pyrkivät mahdollisimman vähällä ponnistuksella sanomaan mahdollisimman paljon. Tulotasoja selitettäessä jakauman selitys on vuorostaan se, että rikkaat suosivat toisia rikkaita (ks. Fagan & Ramazan 2010). Zipfin jakauma ei siis tosiasiassa ole deterministinen seuraantolaki, sillä se ei ole todellisuuteen vaikuttava voima tai periaate, jonka avulla voidaan kategorisesti selittää ilmiöitä. Esimerkiksi termodynamiikan kolmas laki, joka muun muassa lausuu, että eristetyssä systeemissä kaikki erilämpöiset kappaleet aikaa myöten asettuvat samaan lämpötilaan, toteutuu aina. Termodynamiikan kolmas laki on luonteeltaan selittävä seuraantolaki (luonnonlaki), sillä kun laki tunnetaan, voidaan selittää muutoksia ilmiössä. Zipfin jakauma vuorostaan kuvaa empiiristä yhteenliittymisen säännönmukaisuutta, joka on löydettävissä niin biologisista kuin

sosiaalisistakin systeemeistä. Jakauma ei kerro, mikä aiheuttaa ilmiöt vaan millä todennäköisyydellä jokin asia toteutuu.

Säännönmukaisuus tarkoittaa sitä, että suuressa otoksessa lähes kaikki ilmiöt noudattavat jotain jakaumaa. Poikkeuksiakin toki löytyy. Zipfin jakaumaa ei esimerkiksi esiinny kiinan kielessä. Koska Zipfin jakauma ei ole deterministinen seuraantolaki, se ei itsessään kerro, miksi se ei aina toteudu. Silloin kun jakaumaa ei esiinny, joudutaan esittämään erillinen selityshypoteesi. Esimerkiksi Dahui, Li Menghui ja Di Zengru (2005) esittävät hypoteesin siitä, että uusien kiinan sanamerkkien luominen ja merkistön satunnainen muuttuminen ovat hitaampia kuin kielissä, joissa on äänteisiin perustuva aakkoskirjoitus.

Realismiin sisältyy usein käsitys, jonka mukaan maailma on lainomainen (noominen) ja on olemassa empiirisiä deterministisiä luonnonlakeja. Tieteen tehtävä on löytää reaalitodellisuuden seuraanto- ja yhteenliittymälait. Tällaisen maailmankuvan houkuttelemat kielentutkijat nimittävät herkästi myös kielitieteessä havaittuja säännönmukaisuuksia *laeiksi* silloinkin, kun on varsinaisesti kysymys tilastollisesta jakaumasta tai muusta säännönmukaisuudesta. Tutkijan tulee siis olla tarkka siitä, millaisia **käsitteitä hän käyttää (**esim. ovatko ne keskenään koherentteja), ja tarkastella, miten käsitteistö ohjaa hänen tutkimuksensa suuntaa ja päätelmiä.

Kausaalisuus sekaantuu tai yhtäläistyy helposti käsitteen seuraantolaki kanssa. Kaikki kausaalisuus eli aiheutus ei johdu deterministisistä luonnonlainomaisista tekijöistä (vrt. esim. normaalissa ilmanpaineessa divetyoksidi eli vesi höyrystyy 100 celsiusasteessa). Esimerkiksi pelästynyt rotta voi kiihottaa muut rotat ympäristössään juoksemaan eri suuntiin. Mentaalinen tila siis aiheuttaa ulkoisesti havaittavaa käytöstä. Vaikka samankaltaisen mentaalisen ilmiön vaikutuksen joukkokäyttäytymiseen ja sen säännönmukaisuuden voi havaita myös muissa eläinlajeissa, kyseessä ei ole seuraantolaki. Sosiaalisesti jaetut ilmiöt aiheuttavat säännönmukaisuuksia populaation tasolla, koska niillä on yksilöiden toimintaa suuntaava ja ohjaava samankaltainen vaikutus – näin on myös kielessä. Jotkin yleistettävät lauseet voivat olla selitysvoimaisia ja myös ennustaa tapahtumia olematta kuitenkaan seuraantolakeja. Esimerkiksi lauseet ”pelästynyt rotta juoksee” ja ”kieli on ihmisen

ominaisuus” voidaan parhaiten ymmärtää teoreettisina kvalitatiivisina yhteenliittymislakeina.

Kielentutkimuksessa ehkä lainomaisimmin ymmärretty käsite on ollut äännelaki. Se on historiallis-vertailevassa, erityisesti nuorgrammaattisessa kielentutkimuksessa käytetty käsite, jolla tarkoitetaan tietyllä ajanjaksolla, tietyssä kieliyhteisössä ja tietyssä äänneympäristössä tapahtunutta säännöllistä äänteenmuutosta (vrt. synkroninen äänteenvaihtelu). Tällainen on ollut esimerkiksi *t*:n assibilaatio *s*:ksi *i*:n edellä (*kä**si**ä* < *kä**te**+i*, missä monikon *i* ottaa vartalovokaalin *e* paikan ja assibiloi sitten vartalon toisen tavun konsonantin *t*). Tämä *ti* > *si* -muutos on pystytty ajoittamaan suhteellisen kronologian avulla kantasuomen sellaiseen jaksoon, jolloin kieleen tuli runsaasti (vanhimpia tunnettuja) balttilaisia lainasanoja. Nämä osallistuivat *ti* > *si* -muutokseen (esim. *silta* < liett. *tiltas* ’silta’) kielen omaperäisen sanaston ohella. Sen sijaan myöhemmät lainat, esimerkiksi germaaniset (esim. *äi**ti***) tai slaavilaiset (esim. *ka**ti**ska*) lainat eivät osallistuneet muutokseen, sillä äännelaki ei ollut tuolloin enää vallitseva. (Äännelaista ks. Historiallinen ja vertaileva kielentutkimus tk.) Kyseessä ei kuitenkaan ole empiiristä seuraantolakia vastaava välttämätön voima, joka deterministisesti ohjaisi kieltä. Äännelaki muodostuu joukosta ehtoja, joiden täyttyessä äännejärjestelmässä voi tapahtua muutoksia. Lainomaisen siitä tekee vahva vaikuttavuus sen voimassaoloaikana ehtojen täyttyessä mutta vastaavasti vähemmän lainomaisen sen ajallinen tilapäisyys. Sen taustalla voi nähdä myös yleisiä periaatteita, kuten ”kaikissa kielissä esiintyy muutoksia” tai ”äänteiden muutos ei kunnioita kieliopillisia jakoja”. Nämä periaatteet eroavat deterministisistä luonnonlaeista, esimerkiksi painovoimasta, joka lausuu, että kaikki massalliset kappaleet vetävät toisiaan puoleensa. Painovoima on universaalisti tosi ja selittävä.

Pelkkä *laki*-sanan käyttö ei siis merkitse, että kielessä vaikuttaisi luonnonlain kaltaisia voimia. Kielen muutoksia rajoittavat periaatteet ovat useimmiten luonteeltaan osittain loogisia ja biologisia (esim. ”mikään äänne, jota ihmisen kuulojärjestelmä ei pysty erittelemään, ei voi olla merkitystä erottava tai välittävä”), tai periaatteiden motivaatio on ihmisen ja ihmisryhmien tavoitteisiin liittyvä eli teleologinen. Kielentutkimuksessa on siten mielekästä pohtia, millä lailla kausaalisuus,

säännönmukaisuus tai loogiset rajoitteet on käsitteellistetty metodin taustateoriassa. Kausaalisuuden ja lainomaisuuden käsitteellinen sekoittaminen toisiinsa voi aiheuttaa erheellisiä mielikuvia kielen luonteesta ja ohjata siten myös tutkimusta väärään suuntaan. Kuten tässä luvussa aiemmin esitetystä taulukosta 1 huomataan, lainomaisia ilmiöitä on erilaisia, ja seuraantolait ovat oma erityinen luokkansa.

Chomskyn olettamat biologista alkuperää olevat **kielen universaalit**, joiden mukaan puhutuilla luonnollisilla kielillä on yhteinen perusrakenne, nostetaan toisinaan myös keskusteluun käsitteen *laki* yhteydessä. Tämä voi kuitenkin johtaa harhaan. Esimerkiksi käsitys siitä, että västäräkin viserrys eroaa tuulihaukan äänestä, koska näiden kahden linnun DNA on erilainen, ei oikeuta päätelmään, että västäräkin viserrys on lainomainen. Västäräkin viserrys, tuulihaukan ääni ja ihmisen luonnollinen kieli ovat lajityypillisiä piirteitä. Geenit sisältävät kontingenttia (ks. Johdannon luku 2.3. tk.) informaatiota, joka sekä mahdollistaa että rajoittaa suurelta osin kaikkien elävien organismien fenotyypin funktioita. Lajikohtaisissa funktioissa on siksi odotettavissa lajin sisäistä säännönmukaisuutta, ja odotettavaa on myös, että osa funktioista muuttuu evoluution myötä. Chomskylaisittain ajatellut kielen universaalit eivät siten ole luonnonlakiin verrattavissa oleva ilmiö, vaan ne ovat ymmärrettävissä lajityypillisenä säännönmukaisena funktiona niiltä osin kuin kielellä on geneettistä perustaa. Loogisesti on myös oleellista huomata, että siitä väitteestä, että jokin kielen piirre on biologian rajoittama tai motivoima, ei seuraa, että kaikki kielen keskeiset elementit olisivat geneettistä alkuperää. Biologinen rajoite kielelle ei myöskään tarkoita, että rajoitteiden täytyy olla kielispesifejä: ihmisen muisti, kognitiiviset prosessit ja havaintomekanismit kaikki asettavat rajoitteita sille, mikä kielessä ja kommunikaatiossa on mahdollista. Kieliopillisten tai semanttisten rakenteiden redusoiminen yksinomaan geneettiseen informaatioon ei siten ole motivoitua, vaikka käsitteeseen *universaali* tällainen elementti voidaan liittää.

Toisessa merkityksessä universaaleista on puhuttu kielitypologiassa. *Typologisilla universaaleilla* tarkoitetaan maailman kieliä ja ihmiskieltä yleensä koskevia havaintoja, jotka perustuvat suurten kielimäärien pohjalta tehtyihin induktiivisiin yleistyksiin. Typologiset universaalit ovat

tyypillisesti vahvoja tilastollisia tendenssejä, ja täysin poikkeuksettomat, absoluuttiset universaalit ovat harvinaisia. Termi *universaali* on näin erilaisessa käytössä typologiassa kuin yleensä tieteenfilosofiassa, jossa tilastollinen ja universaali ovat toistensa vastakohtia. Lakien sijaan tällaiset universaalit edustavat pikemmin siis erivahvuisia tendenssejä kielten osajoukosta. Vaikutelma lainomaisuudesta vahvistuu sitä mukaa, mitä vahvemmasta tendenssistä on kyse. Typologiassa tehdään myös jako implikationaalisiin ja ei-implikationaalisiin universaaleihin. Implikationaaliset universaalit sisältävät ehdon (esim. kaikissa kielissä pätee, että jos kielen perussanajärjestys on SOV, sen adpositiot ovat ensisijaisesti postpositioita eivätkä prepositioita). Ei-implikationaaliset ovat yksinkertaisia väittämiä (esim. kaikissa kielissä on vokaaleja ja konsonantteja). Sekä implikationaaliset että ei-implikationaaliset universaalit voivat olla poikkeuksettomia absoluuttisia universaaleja tai vahvoja tilastollisia tendenssejä. Typologisia universaaleja on esitetty kymmeniä. (ks. Kielitypologian menetelmät tk.).

Laeista puhuttaessa voi vielä mainita Gricen maksiimit eli vuorovaikutuksen yhteistyöperiaatteet (Grice 1975). Näitä ovat laadun (”puhu totta”), määrän (”ole riittävän informatiivinen”), relevanssin (”pysy asiassa”) ja tavan (”ole ymmärrettävä”) maksiimit. Vaikka Grice tarkoitti maksiiminsa deskriptiivisiksi luonnehdinnoiksi vuorovaikutuksen yleisistä periaatteista tai ideaalista, niitä on usein tulkittu lainomaisina preskriptiivisen normituksen merkityksessä. Kyse on kuitenkin hyvin yleisen tason sosiaalipsykologisista periaatteista, jotka ovat paljolti kulttuurisesti määräytyneitä tavanomaisuuksia. Eri kulttuuripiireissä maksiimit voivat korostua eri tavoin. Tässä *laki*-termi on käytössä velvoittavan kulttuurisen säädöksen merkityksessä, mutta maksiimeja voisi luonnehtia määrällisiksi tai laadullisiksi tilastollisiksi yhteenliittymislaeiksi: relevanssista poikkeavat keskustelupuheenvuorot ovat normaalisti jakautuneita. Järkevintä tutkimuksessa on kuitenkin olla käyttämättä laki-käsitettä silloin, kun ei ole kyse deterministisestä laista. Läpinäkyvä käsitteellinen selitys yhteenliittymisen tai tilastollisuuden laadusta on tieteen etiikan mukainen rehellinen esitys.

Erilaisia selittämismalleja on lukuisia. Edellä esitelty Zipfin jakauma ilmentää tilastollista selittämistä, ja kielitypologian esittämiin

universaaleihin pyritään löytämään funktionaalisia selityksiä. Äännelait kuvaavat historiallista tai evolutionaarista selittämistä. Gricen maksiimeja voi pitää vuorovaikutuksen järjestymisen funktionaalisena selittämisenä. Samoin analogiaa äännelakien vastinparina voi pitää kielisysteemin jäsentymisen funktionaalisena selittämisenä. Muita selitysmalleja ovat muun muassa kausaalinen, intentionaalis-teleologinen (tavoitteisiin ja päämääriin perustuva) ja geneettinen (alkuperään tai syntyyn perustuva) selittäminen.

Selittämisestä voidaan pyrkiä vielä edemmäs asioiden ja ilmiöiden ennakoimiseen ja ennustamiseen, jolloin ihminen voi toimintaansa järjestelemällä tavoitella haluamiaan tuloksia. Esimerkiksi kun lapsen viivästynyt puheenkehitys tunnistetaan (ilmiön kuvaus) ja tilan aiheuttanut tekijä ymmärretään (esim. geneettinen selitys, fysiologinen tekijä), on mahdollista harjoittaa kieltä niin, että viivästyksen aiheuttama heikennys lapsen elämänlaadulle vähenee tai poistuu kokonaan. Tällaisella soveltavalla kielentutkimuksella voidaan vaikuttaa takaisin maailmaan, mutta se edellyttää pätevää kuvauksen ja selittämisen perinnettä.

Selittäminen tieteessä yleensä tai kielentutkimuksessa erityisesti (esim. Wunderlich 1979 [1974], luku 3) kattaa ison ja haastavan tieteenfilosofian osa-alueen, jota on tutkittu valitettavan vähän. Suomalaisessa kielentutkimuksessa funktionaalista selittämistä kielen muutoksen kuvaamisessa on käsitellyt Määttä (1994). Konkreettinen ja havainnollinen esitys selittämisestä on Wiikin (1984) tutkimus, jossa hän esittää lukuisia tapoja selittää suomen taivutusjärjestelmän sellaista erikoisuutta, että voidaan ja pitääkin sanoa *kana* : *kanoja* mutta ei *muna* : **munoja* (vaan *munia*).

Edellä on huomattu läheinen suhde todellisuuskäsityksen ja kielentutkimuksen tutkimuskäsitteistön välillä. Samalla on hyvä panna merkille myös erilaisten tiedekäsitysten ja -ihanteiden vaikutus tieteessä yleensä ja tieteenalojen kesken. Lisätutkimus kielentutkimuksen käsitteistöstä ja erityisesti loogisia ristiriitoja aiheuttavista termeistä tutkimuksessa olisi tervetullutta.

### 3.2.3. KIELEN REIFIKAATIO, NORMI JA ONTOLOGIA

*Reifikaatioksi* kutsutaan abstraktin ilmiön muuttamista konkreettiseksi. Metodologiassa sillä tarkoitetaan, että konstruoidulle tutkimuskohteelle osoitetaan ontologinen asemapaikka. Tutkija esimerkiksi päättää kutsua aineistosta noussutta ilmiötä nimellä X (esim. *kongruenssi*). X on termillä nimetty teoreettinen konstruktio, jonka määritelmä muodostuu toisista käsitteistä ja viittauksista aineiston ominaisuuksiin. *Konstruktiovaliditeetti* on myös metodologiassa käytetty käsite, joka tarkoittaa sitä, missä määrin tutkimuksen muuttujat mittaavat täsmällisesti teoreettista konstruktiota (Vogt 1999, 53). Reifikaatio on siis tavallaan käsitteellinen oikopolku, jossa ei lopulta ole huomioitu kaikkia niitä tutkimuksen vaiheita, jotka johtivat teoreettisen konstruktion luomiseen (datan käsittely, transkribointi, analyysi ja luokittelu jne.). Sen sijaan nimeämällä luodaan diskreetti olemassaolo teoreettiselle yläkäsitteelle. Reifikaation asteita on useita, ja ontologisoiminen on niistä vahvin.

Jo arkinen kieli tarjoaa yhdenlaisen käsityksen kielestä reifioituna: ”Mitähän kieltä tuo puhuu? Millainen kieli on chickasaw? Kuinkahan helposti sen kielen oppii?” Arkinen kielenkäyttömme ohjaa ajattelemaan kieltä rajautuvana asiana. Substantiivi *kieli* nimeää olion, joka on abstrakti mutta kuitenkin ikään kuin rajoiltaan hahmottuva. Näin se näyttäytyy enemmän esinemäiseksi käsitteellistettynä kuin toimintana, jota esimerkiksi kielen tavalliset ilmentymät, kuten puhuminen, kuunteleminen, väittäminen ja lupaaminen, ykskantaan ovat (ks. Johdannon luku 3.1. tk.). Kieli voi siis hahmottua oliona – toki erilaisena oliona kuin konkreettinen kivenmurikka tai naapurin Matti.

De Saussuren (2002 [1916]) merkkikäsitys on myös yksi esimerkki reifikaatiosta. Kielellinen merkki on konkreettinen objekti: ”The signs comprising a language are not abstractions, but real objects. Linguistics studies these objects and relations between them. They may be termed the concrete entities of that science.” (Mts. 144.) Vaikka de Saussure (mt.) onkin kielen teorian klassikko, niin on olemassa useita tulkintoja siitä, mitä hän tarkalleen tarkoitti.[3]

Metafysiikan perusteorioihin tutustuminen on kielentutkijallekin tarpeellista. Vaikka erityisesti puhutun kielimuodon raakadatasta voi olla vaikea löytää itsenäisiä, autonomisia ja diskreettejä yksiköitä, usein

kysymykseen, mitä kieli on, vastataan aivan kuin tällaisia itsenäisiä yksiköitä olisi olemassa. Tämä ilmentää tapaa tavoitella yleistyksiä kielestä. Esimerkiksi käsitys kielestä sosiaalisena normistona on yksi ominainen tapa käsitteellistää kielen olemusta. Termit *normi* ja *normisto* viittaavat käsitykseen, että kieltä voidaan koettaa käsitteellistää normien diskreettinä osajoukkona L $\{n_1 ... n_n\}$. Sosiaalinen normisto ei kuitenkaan ole ainoa tapa kuvata kielen olemusta. Generatiivinen kielentutkimus olettaa, että kieli on yhtä kuin kielikyky: yksittäisten kielten säännöt ovat seurauksia sisäsyntyisestä universaalista kieliopista (esim. Chomsky 2007, 50–52; Brattico 2008, 9). Kielen normeja ei sisäistetä niinkään oppimisen myötä, vaan kielen ulkoisia säännöstöjä kehittyy sisäisen kielikyvyn seurauksena. Tämän käsityksen mukaan kieli on sellainen sisäsyntyisten sääntöjen osajoukko, joka generoi kaikki kieliopilliset lauseet L $\{s_1 ... s_n\}$. Kieli normistona tai universaalina kielioppina ovat ilmentymiä pyrkimyksestä kielijärjestelmän reifioimiseen.

Viimeistään de Saussuren kielikäsitys johti kysymykseen kielen ja sen yksiköiden identiteetistä ja ontologiasta (de Saussure 2002 [1916], 154). Sama kysymys askarrutti myös Wittgensteinia (1997 [1953], 558), ja sekä de Saussure että Wittgenstein antoivat lopulta samankaltaisen vastauksen: yksikön merkitys rakentuu muita kielellisiä yksiköitä vasten. Vain osana kielijärjestelmää voi yksiköllä (äänteellä, morfeemilla, sanalla ym.) olla merkitys (Harris 1988, 19). Kielet ovat järjestelmiä eli joidenkin periaatteiden toiminnallisia kokonaisuuksia (Wittgenstein 1997 [1953], 170).

Käsitykset siitä, miten kielijärjestelmät ovat olemassa, perustuvat kuitenkin erilaisille premisseille: De Saussure perustaa merkit niiden keskinäisten assosiaatioiden varaan, Wittgenstein sääntöjen seuraamiseen. Wittgensteinin mukaan sääntö on olemassa ainoastaan sen seuraamisen kautta, eikä sillä siten ole itsenäistä olemassaoloa yksilön mielessä. Kielisääntöjen seuraaminen on konkreettinen aktiviteetti (ks. Baker & Hacker 1988, 154–156). Wittgenstein kieltää mentaalisuuden kielijärjestelmän perustana: kieli ei saa merkitystä merkkien välisten assosiaatioiden kautta.

De Saussure ja Wittgenstein tarjoavat siis kaksi keskenään erilaista käsitystä kielestä järjestelmänä. Kumpikaan ei ole normisto. Termi

*normi* tulee latinan sanasta *norma*, joka alun perin tarkoitti puusepän työkalua, jolla tehtiin tasaisia muotoja ja kulmia. Muotilla siis standardisoitiin puutyöt. *Normi*-termin käyttö kielitieteellisissä tutkimuksissa yleistyi vasta 1950- ja 1960-luvuilla etno- ja sosiolingvistiikassa. Sosiolingvistiikassa Labov ei käyttänyt *normi*-termiä kuitenkaan vielä 1970-luvulla vaan käsitteellisti yleensä ihmisen kielellistä toimintaa ohjaavan laskennallisen ilmiön muuttujasääntönä (engl. *variable rule*). On huomattava, että termiä *normi* käytettiin ja käytetään monesti edelleen näillä aloilla ulkoisen käyttäytymisen selittäjänä eli esimerkiksi merkityksessä 'normi on odotettavissa oleva ryhmän toiminnan malli' tai 'normi on vuorovaikutuksen muoto, jonka välityksellä yhteisö saa jäsenensä toimimaan' (ks. myös Crystal 2008 s.v. *norm*). Normi on siis näin määritellen aineistoperustaisen tutkimuksen havaintoja systematisoiva konstruktio, jota käytetään ihmisten ulkoisen kielikäytöksen kuvaamiseen (ks. kuitenkin Sherif 1965). *Normi*-termin merkityksestä on kielitieteessä keskusteltu laajasti 1970-luvulta alkaen (ks. esim. Kauhanen 2006; Itkonen 2019).

Käsitys sosiaalisista kielellisistä normeista sisäistettynä järjestelmänä ei vaikuta olevan osa kielen mentaalisia prosesseja tutkivien alojen käsitejärjestelmiä. Kielipsykologian, psykolingvistiikan tai neuropsykologian tutkimuksista, perusteoksista ja ensyklopedioista ei yleensä löydy termiä *normi*. Nämä tutkimusalat tarkastelevat ihmisen aivojen ja mielen sisäisiä kielen ymmärtämisen ja tuottamisen prosesseja. Edes kognitiivisen kielitieteen perusteoksien hakemistoista ei tunnu löytyvän termiä *normi* sellaisenaan (ks. Langacker 1987; Lakoff 1987; Tomasello 2005; vrt. Mäkilähde & Hynönen 2019).

Kielen sosiaalista ontologiaa, sosiaalista normistoa, ihmismielestä erillisenä ja itsenäisesti olemassa olevana järjestelmänä ovat ehdottaneet Searle (2008) ja Itkonen (1997). Searlen (mt.) näkemyksen mukaan kieli on instituutio, joka mahdollistaa muiden sosiaalisten instituutioiden, kuten talous- ja sopimusjärjestelmän, valtion ja muiden sosiaalisten sopimusten ja yhteisliittymien, olemassaolon. Itkonen (1997; 2019) vuorostaan perustaa kielen sosiaalisen ontologian normatiivisuuteen. Kielen sosiaalista puolta korostavilla filosofioilla on myös taipumus reifikoida ja ontologisoida teoreettisia käsitteitä. Esimerkiksi sellaiset luonnehdinnat kuin "language is a social entity" tai "language is an object of

common knowledge" ilmaisevat käsityksen kielestä objektina ja oliona. Reifikaation vaikutus kielen metateoriaan on silloin ilmeinen, kun oletetaan, että jokin on objekti, olio tai entiteetti. Tällöin on myös vastattava metafyysisiin kysymyksiin, missä ja miten tuo olio on olemassa.

Tutkijoiden edustamia erilaisia kielikäsityksiä voi sijoittaa Popperin (1978) kolmen maailman malliin (ks. Johdannon luku 3.2.1. tk.). Taulukossa 2 joitakin edellä mainittuja kielen teorioita on sovitettu Popperin kolmen maailman rakenteeseen.

| MAAILMA | NORMI | MUU |
|---|---|---|
| Maailma 1 (M1): fysikaalinen taso | Labov (?) | Wittgenstein (sääntö) |
| Maailma 2 (M2): mentaalinen taso | | Langacker (skeema), Chomsky (universaalikielioppi) |
| Maailma 3 (M3): sosiaalinen taso | Searle | |

Taulukko 2. Popperin maailmat ja erilaisia kielikäsityksiä.

Taulukossa 2 Labovin ajattelu on varauksella sijoitettu normikategoriaan, sillä kuten Kauhanen (2006) on havainnollistanut, Labov aloitti kielitieteessä uudenlaisen keskustelun juuri normien luonteesta, vaikkei aina tuota käsitettä käyttänytkään. Labov kohdisti normin tarkastelun paikallisen kieliyhteisön käyttäytymiseen, jota voi havaita empiirisesti. Keskiössä oli kieliyhteisön muodostuminen. Tämä oli ristiriidassa tuolloin dominoivan generativistisen kielinäkemyksen kanssa, jossa normisysteemi oli abstrakti järjestelmä, johon oli pääsy intuition (kielitajun) kautta.

Wittgensteinin filosofia sijoittuu maailmaan 1. Hänen esittämänsä yksityiskieliargumentti on suunnattu kielen psykologismia vastaan: sanat eivät saa merkitystään puhujan sisäisistä assosiaatioista, muiden tavoittamattomissa olevista yksityisistä mielentiloista. Yksityinen kieli on vain yhden ihmisen ymmärtämä, sillä se perustuu ainutlaatuisiin jakamattomiin mielensisältöihin. Niin de Saussure kuin Wittgensteinkin korostivat kielen sosiaalista puolta, mutta on hyvä tiedostaa, että ei ole olemassa suoraa yhteyttä konkreettisen M1:n ja sosiaalisen M3:n välillä

ilman ihmisen mieltä (M2). Wittgensteinin käsitys säännön seuraamisesta on ulkoista seuraamista, samoin kuin Labovilla.

Popperin kolmen maailman malli perustuu materialismiin, eli se olettaa, että kysymys materian (M1) ontologiasta on ongelmaton ja ainoastaan tietoisuus (M2) ja sosiaalinen taso (M3) tarvitsevat perusteluja. Tämä johtaa siihen, että jos olettaa jonkin materiasta erillisen mentaalisen tai sosiaalisen todellisuuden, jossa normit tai säännöt sijaitsevat, on vastattava kysymykseen, mikä on se laki tai periaate, joka selittää sen, miten nämä erilaiset substanssit ovat keskenään tekemisissä. Jos normi tai mieli voi vaikuttaa materiaan, niin silloin se on samassa todellisuudessa materian kanssa ja asetelma on monistinen.

Emergenttinen materialismi olettaa, että materiasta emergoituu mieli ja monien mielten yhteistoiminnoista kieli. 1990-luvulla kannatusta saanut supervenienssi-käsite (engl. *supervenience*, suom. myös *päältävyys*) sanoo, että kaksi ilmiötä, esimerkiksi neurologinen aivotoiminta ja tietoisuus, toimivat samaan aikaan. Ilman aktiivisuutta yhdessä ei ole aktiivisuutta toisessa. Tämä on kuitenkin analogisesti sama asia kuin sanoa, että Matissa ja Itämeressä tapahtuu muutos samanaikaisesti. Väite samanaikaisuudesta ei selitä kahden ilmiön yhteyttä, eikä se ole vastaus kysymykseen, miten fysikaalinen ja mentaalinen vaikuttavat toisiinsa.

Myös materialismia itseään kohtaan voi suunnata kritiikkiä esimerkiksi Searlen (1980) kiinalaisen huoneen tapaisilla analogioilla: Suljetussa huoneessa on mies, joka ei osaa kiinaa, mutta hänellä on sääntökirja, joka kertoo, mikä piktogrammi tulee liittää toiseen. Luukusta tulee merkki, johon mies vastaa toisella merkillä. Hän toimii siis tilanteessa sääntökirjan ohjeen mukaan ja siten täysin korrektisti, ja huoneen ulkopuolella olevat uskovat miehen osaavan kiinaa. Analogiaa voidaan laajentaa uusiin teknologioihin: Jos kieli ja tietoisuus eivät ole mitään muuta kuin materiaalisia prosesseja, ymmärtääkö kielenkorjausohjelma kieltä? Entä puhuuko kieltä ihminen, joka toistaa mekaanisesti vieraan kielen sanoja Google-kääntäjästä? Ovatko tällaiset mekaaniset toiminnot samoja intuitioita tai toimintoja kuin ihmisen kielen normia koskevat kieli-intuitiot (kielitaju) ja puhe?

1990-luvun lopulta lähtien erilaiset systeemiset filosofiat ovat esittäneet sen näkemyksen, että materiaalin, mielen ja sosiaalisen ongelma

olisi väärän kysymyksenasettelun tulos. Esimerkiksi **rakennerealismi** esittää, että kaikki (niin materia, mieli kuin sosiaalinen) koostuu relaatioista, rakenteista ja systeemeistä. Olioilla ei ole sinänsä itsenäistä olemassaoloa. Kieli tämän näkemyksen mukaan on useiden systeemien summa eli itsenäisten sääntöjen ohjaamien toimivien osien muodostama kompleksinen järjestelmä, jonka toimintasäännöt (normit, säännöt) muodostuvat osajärjestelmien välisestä vuorovaikutuksesta. Edellä luvussa 3.2.1. esitetyn kuvion 17 laatikot symboloivat systeemejä. Systeemien interaktiivisista relaatioista syntyy uusia dynaamisia relaatioita, jotka vaikuttavat takaisin systeemeihin. Esimerkiksi muutokset kielenkäytön eri tyyleissä ja rekistereissä vaikuttavat muihin systeemeihin. Systeemiset relaatiot muuttuvat ajassa, ja uusia ilmiöitä syntyy koko ajan. Kieltä ei siis voi rakennerealismin sisäisesti redusoida vain kolmeen ontologiaan. Kieli on systeemisten relaatioiden dynaaminen, muuttuva ja alati kehittyvä joukko, jota ei ehkä voi laisinkaan tyhjentävästi määritellä.

Rakennerealismin keskeinen ajatus on, että eksistenssi on relaatiossa olemista. Tämä tarkoittaa, että erilaisia olemisen tapoja on periaatteessa yhtä paljon kuin voidaan kuvitella erilaisia relaatioita. Rakennerealismia kutsutaan siksi myös **sademetsärealismiksi** kielikuvana sademetsän monimuotoisuudesta ja runsaslajisuudesta. Kielitieteen metateorian kannalta ajatus on mielenkiintoinen. Se tarkoittaa, että rakennerealismissa ei tarvitse sitoutua esimerkiksi vain yhteen normin tai säännön määritelmään. Oletus on, että normeja on erilaisia ja että kielen sääntöjä muodostuu eri tavoin. Tämä myös tarkoittaa sitä, että kielitieteen eri tutkimusnäkökulmien tutkimusten tulokset voivat periaatteessa kaikki olla valideja. Ne voivat kertoa tosia asioita kompleksisen systeemin toiminnoista.

Edellä on huomattu, että sellainen keskeinen ja perustava kielentutkimuksen käsite kuin *normi* johtaa monimutkaisiin ontologisiin kysymyksiin ja metafyysisiin sitoumuksiin. Tutkijan tulee siis pohtia reifikaatiota ja sitä, millaisen olemassaolon hän osoittaa jollekin kielentutkimuksen teoreettisesti konstruoidulle käsitteelle ja millä perustein. Tämä määritelmä nimittäin sitoo tutkimuksen johonkin ontologiaan.

## 3.3. Kielentutkimuksen olemuksesta

### 3.3.1. KUVAUKSEN SUHDE KOHTEESEEN: ESIMERKKEINÄ FORMALISMI JA FUNKTIONALISMI

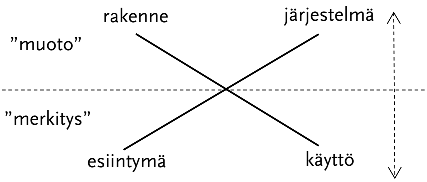
Tässä luvussa tarkastellaan sitä, millä tavoin suhde tutkimuksen kohteeseen eli kieleen rakentaa tutkimusohjelmaa. Erilaisten tutkimusohjelmien jaottelut (esim. Scholz, Pelletier & Pullum 2016; ks. Johdannon luku 3.2. tk.), uskomukset kielen luonteesta ja arvopainotteiset premissit tutkimuksen tavoitteista ohjaavat kielentutkimusta. Tässä avataan näkökulmia kielitieteelliseen tutkimukseen alalla melko vakiintuneen formalismi–funktionalismi-jaon avulla. Tavoitteena on tarkastella, miten kielentutkimuksen kohde, luonnollinen kieli (puhuttu, viitottu tai kirjoitettu), on aina rajauksen ja näkökulman vanki.

Luonnollinen kieli voidaan rajata nimeämällä kielen jokin muoto esimerkiksi epämuodolliseksi arkipuheeksi, juhlapuheeksi, vihapuheeksi tai huolitelluksi yleispuhekieleksi. Eri luonnehdinnat synnyttävät käsityksiä, millaisesta kielestä ja kielellisestä toiminnasta on kulloinkin kysymys. Kielentutkimuksessa myös nojataan usein oletukseen siitä, että kuvattu ilmiö kuuluu kieleen tai johonkin sen muotoon, varieteettiin (lektiin).[4] Esimerkiksi korpuksiin perustuvan tutkimuksen yksi lähtökohta on, että käytettävä korpus edustaa jotain kieltä tai sen osaa (esim. suomenkielistä sanomalehtitekstiä 2010-luvulla). Kielentutkimuksessa on siten kyseessä aina jokin näkökulmainen rajaus kieleen: tavoitteena on tehdä yleistyksiä kielestä (esim. johonkin tilastolliseen otantaan perustuvan) aineiston perusteella. Sitä, millainen olemus tällä yleistyksen kohteella eli kielellä on, käsitellään seuraavaksi.

Kuvio 18 ilmentää karkeasti kahden keskeisen kielitieteen koulukunnan tai katsantokannan, formalismin ja funktionalismin, suhtautumista kielen muotoon ja merkitykseen. Formalistinen suuntaus (ks. esim. Chomsky 2001 [1965]) keskittyy kuvaamaan kielen rakenteita ja niiden muodostamaa kielen järjestelmää. Näkemys esittää toisinaan, että kielen muoto on täysin autonominen (1) (ks. kuvio 18). Sen mukaan muotoon ei siis liity merkitystä, eikä merkityksen kuvauksen välttämättä katsota kuuluvan lainkaan kielentutkimuksen alaan (esim. Bloomfield 1957 [1933], luku 9). Astetta lievemmässä kannassa oletetaan, että kielen

muoto on pääasiallisesti autonomista mutta käyttö vaikuttaa paikoin muotoon (2). Tämän voi tulkita joko niin, että käyttö vaikuttaa muodon ilmiasuun (esim. runsas käyttö kuluttaa muotoa), tai niin, että käyttö koodaa jonkin verran merkitystä.

Funktionalistinen suuntaus taas keskittyy usein kuvaamaan kielen käyttöä havaittavien esiintymien kannalta (ks. esim. Haddington & Sivonen 2010). Se kuvaa kieltä erityisesti merkityksen muodostumisen ja välittymisen valossa silloin, kun kieli on osa sosiaalista toimintaa. Kuviossa 18 funktionaalisista näkemyksistä lievemmän mukaan kielen muodon ajatellaan perustuvan pääsääntöisesti käyttöön, mutta poikkeuksiakin on (3). Tällöin merkitystä ajatellaan ennen kaikkea pragmatiikkana ('merkitys on käyttöä'), mutta osa merkityksestä nähdään vakiintuneena. Samalla kielen muoto voidaan nähdä vahvasti käytön motivoimana. Hyvin funktionalistinen näkemys olettaa, että kielen muoto perustuu vain käyttöön (4). Muotoa ei tavallaan ole ilman sen konkreettista ilmentymää käytössä, ja merkitys on hyvin pragmatistista; se kehkeytyy vain kielenkäyttötilanteissa ajallis-paikallisesti ja joskus paljolti ilman kielellistä muotoa muin viestinnällisin keinoin (esim. elein ja ilmein).



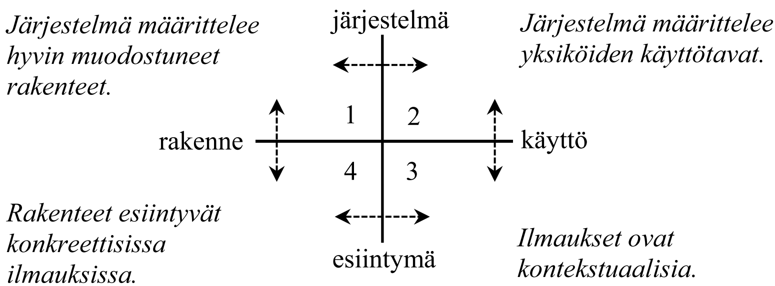

1. Kielen muoto on autonominen.
2. Kielen muoto on pääasiassa autonominen, mutta joiltain osin käyttö vaikuttaa muotoon.
3. Poikkeuksia lukuun ottamatta kielen muoto perustuu olennaisesti käyttöön.
4. Kielen muoto on täysin käytöstä riippuvaa.

Kuvio 18. Formalistisen ja funktionalistisen suuntautumisen taustaoletusten eroja (soveltaen Nikannetta 2015; 2018, 9–10).

Kuvion 18 karkeistus havainnollistaa käytännössä sitä, mikä näkökulma tutkimuksella on kielen muotoon ja merkitykseen ja pidetäänkö jompaakumpaa ensisijaisena. Toinen näkökanta muotoon ja merkitykseen koskee sitä, oletetaanko kielen käytön taustalle jonkinlainen kielijärjestelmä ja käytön ja systeemin välille vastavuoroinen suhde. Sitä käsitystä, että

kielen käytön taustalla on oma symbolinen (platoninen) todellisuutensa, joka ohjaa kielenkäyttöä, kutsutaan **symbolirealismiksi**. Se eroaa formalistien mentalistisesta kannasta, jossa kielioppi siis nähdään viime kädessä osana ihmisen kognitiivisia prosesseja ja generatiivisessa kielentutkimuksessa vieläpä palautuvan osaksi biologiaa (perimää).

Kielentutkimusta voi hahmottaa erilaisten heurististen mallien kautta. Kuviossa 19 kieltä jäsennetään vertikaalisesti abstraktisuuden ja holistisuuden avulla (vrt. kuvioon 18). Horisontaalinen suunta rakenne–käyttö vastaa taas karkeasti jakoa muotoon ja merkitykseen. Molemmilla akseleilla voi liikkua myös kohtisuoraan (ks. kuvion 19 nuolet): Horisontaalisella rakenne–käyttö-akselilla liike vertikaalisesti tarkoittaa käytön osalta merkityksen tarkastelua joko semantiikkana (kenttä 2) tai pragmatiikkana (kenttä 3). Vastaavasti rakennetta voi tarkastella rakenneskeemoina (kenttä 1) tai konkreettisina rakenneskeemojen instansseina eli esiintyminä (kenttä 4). Samalla tavoin vertikaalisella järjestelmä–esiintymä-akselilla liike horisontaalisesti tarkoittaa järjestelmän kuvausta joko formaalina (hyvämuotoisten) rakenteiden systeeminä (kenttä 1) tai konventionaalistuneena semantiikkana (kenttä 2, esim. kielen muodon vakiinnuttamana käsiterakenteiden järjestelmänä). Yhtä lailla esiintymien kuvaus voi keskittyä niiden kontekstuaalisen käyttöfunktion kuvaamiseen (kenttä 3) tai siihen, miten esiintymät ilmentävät kielen rakennepotentiaalia, esimerkiksi morfologista variaatiota (kenttä 4).




Kuvio 19. Kielen rakenteen, esiintymän, käytön ja järjestelmän nelikenttä (soveltaen Nikannetta 2015; 2018, 9–10; ks. myös kuvio 18).

Nelikentässä kentät 1 ja 2 edustavat järjestelmälähtöistä kielenkuvausta ja kentät 3 ja 4 esiintymälähtöistä kuvausta. Kenttä 1 koskee järjestelmälähtöistä muotoperustaista formalismia. Perustyyppinen tätä edustava tutkimus on generatiivinen kielenkuvaus (erit. varhainen transformatiivis-generatiivinen kielentutkimus) ja sitä vastaavat kielenkuvaukset, joiden keskeinen tavoite on esittää sääntöformalismi vain kielen hyvämuotoisille ja kieliopillisille rakenteille. Semantiikalla on sivuosa tällaisessa kuvauksessa toisin kuin kentän 2 järjestelmälähtöisessä funktionalistisessa lähestymistavassa, jossa kieliopillinen kuvaus on juuri ennen kaikkea semantiikan kuvausta (esim. kognitiivinen kielioppi). Kenttä 3 edustaa esiintymälähtöistä funktionalistista lähestymistapaa, jossa ilmauksia tarkastellaan esiintyminä konkreettisissa käyttöyhteyksissään. Ilmaukset ovat kontekstissaan, ja samalla ne itsessään luovat ja rakentavat kontekstia. Kieltä tarkastellaan ennen kaikkea vuorovaikutuksellisena toimintana ja usein yhtenä osana muuta laajempaa sosiaalista toimintaa. Tällaista tutkimusta edustavat muun muassa keskustelunanalyysi, vuorovaikutuslingvistiikka, lingvistinen antropologia, lingvistinen etnografia ja tekstin- ja diskurssintutkimus (ks. ko. lukuja tk.). Kentälle 4 ei löydy suoranaista tutkimusalaa, mutta esimerkiksi puhutun kieliaineiston litteroimisessa tai transkriboimisessa kielen rakenteet saavat kirjallisen ilmiasun konkreettisina esiintyminä. Tämä on ennen kaikkea metodista toimintaa.

Yleisesti ottaen kielentutkimuksen sisällä hajaannusta ja koulukuntaeroja aiheuttaa sekä rikastuttavaa tieteenfilosofista keskustelua rikkoo oletus siitä, että erilaisten metodien tai teorioiden premissit lähtemättömästi asettaisivat eri tutkimusnäkökulmat toistensa kanssa poissulkevaan asemaan. Esimerkiksi hyväksymällä kognitiivisen kielentutkimuksen keskeiset premissit sulkisi generatiivisen perinteen kokonaan pois tai hyväksymällä vuorovaikutuslingvistiikan premissit sulkisi kielitajun käytön kokonaan pois. Juuri tällaisten metodologisten ajattelutapojen havainnoiminen ja kyseenalaistaminen on kielentutkimuksen tieteenfilosofian kannalta erittäin hedelmällistä.

### 3.3.2. Metodien suhteista: intuitio ja empiria

Tieteessä tehdään toisinaan arvottavia eroja erilaisten metodien välille. Esimerkiksi kielentutkimuksessa kielitajun varaisen ja empiirisiin metodeihin nojautuvan tutkimuksen välille synnytetään eroja, jotka saavat helposti koulukuntamaisia piirteitä puolesta ja vastaan (ks. esim. Phillips 2009).

Kielitaju kuitenkin liittyy monella tavalla myös kielitieteen muihin metodeihin. Ensinnä monissa tutkimusasetelmissa (esim. psyko- ja neurolingvistiset kokeet, ks. Psyko- ja neurolingvistiikka; Kokeellinen fonetiikka tk.) kielellinen syöte on usein tutkijan kielitajullaan laatimaa. Kyselytutkimuksissa tutkijan laatimat kysymykset nojaavat usein kyselyn laatineen tutkijan kielitajuun ja ainakin vastaukset kokonaisuutena edustavat kielenkäyttäjien eräänlaista kollektiivista kielitajua (ks. Tummers, Heylen & Geeraerts 2005, 229). Myös elisitoinnissa voidaan konsultoida suoraan kielenpuhujan kielitajua.[5] Kielitaju on siis oikeastaan välttämättömänä osatekijänä monessa kielentutkimuksen tutkimusmetodissa, ja metodien suhde onkin siksi toisiaan täydentävä. Kun metodien käytännöllinen suorittaminen vaiheistetaan, kielitajulla on useimmissa metodeissa rooli joko tutkimuksen suunnitteluvaiheessa tai tulosten tulkinnassa tai molemmissa. Vain nk. autonominen kielentutkimus nojaa myös aineistonhankinnassa kokonaisuudessaan kielitajuun.

*Kielitajun* käsite problematisoitui kielentutkimuksessa oikeastaan vasta 1960-luvulla sen jälkeen, kun Chomsky (2001 [1965], 2) esitti kieli-intuition metodologiset perusteet. Tuohon asti kysymys siitä, miten kielitieteilijä tunnistaa kielen yksiköitä, ei ollut metodologisesti relevantti tai edes tunnistettu. Hieman samaan tapaan kuin kysymys siitä, miten biologi tunnistaa bakteerin tai fyysikko osaa laskea, ei ollut tutkimusraportissa millään tavalla tarpeellista perustella, kuinka kielentutkijan esittämät kieliesimerkit ovat legitiimejä. Tutkimuskohdetta koskeva argumentaatio muodosti tieteellisen tutkimuksen ytimen.

Kielitajua on pohdittu kielitieteen metateoriassa jonkin verran muttei vielä riittävästi (ks. Konstenius 2014, 142–152). Schütze (2016) esittää erään komputationaalisen teorian. Hän on pohtinut sitä, ovatko kielellisen rakenteen tai ilmaisun kieliopillisuuden (engl. *grammaticality*

*judgement*) ja hyväksyttävyyden arviointi (engl. *acceptability judgement*) samoja asioita sekä onko intuitiolla ja introspektiolla käsitteellisiä eroja. Kielitajun heikkoutena on pidetty sitä, että se voi olla epäluotettavaa ja vaihdella eri puhujien kesken. Se on viime kädessä yksilökohtaista ja solipsististista[6].

Chomsky (2001 [1965]) esitti, että kielitaju on ainoa oikea kielitieteen metodi. Kielitajun roolia ja metodologista oikeutusta on kritisoitu erityisesti empiirisesti suuntautuneiden koulukuntien ja tutkimussuuntien puolelta. Vastaavasti empiiristen tutkimusten heikkoutena on aineiston tietynlainen rajallisuus. Esimerkiksi korpusten heikkoutena on pidetty sitä, että ne sisältävät vain niitä kielenilmiöitä, joista korpukset koostuvat. Tällöin on mahdollista, että esimerkiksi jotain harvinaista mutta epäämättä kieleen kuuluvaa ilmiötä ei löydy korpuksesta ollenkaan. Lisäksi myös puhutut kielelliset tilanteet sellaisinaan ovat kertaluontoisia, ne tavallaan edustavat vain itseään. Tämän kokoelman artikkeleissa on pyritty nostamaan esiin myös metodien erilaisia rajoitteita niin, että tutkijan laatima yleistys on rajattu aineiston ja analyysin puitteissa ekologiseen validiteettiin[7].

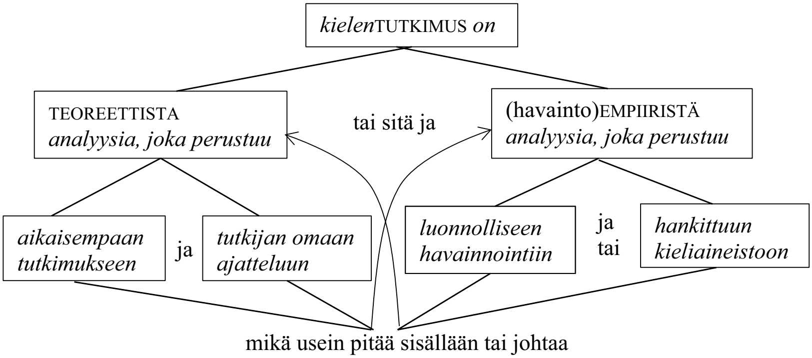
Korpustutkimusten suuntautumistapojen esitetään usein perustuvan kielitajun heikentyvään ja korpusaineiston vahvistuvaan rooliin tutkimuksessa. Aineistoesimerkein tuetussa (engl. *corpus-illustrated* tai *corpus-exemplified*) tutkimuksessa tutkijaa ohjaa ennen kaikkea kielitaju ja teoreettisen viitekehyksen synnyttämät kysymykset. Korpusesimerkit toimivat lähinnä havaintoaineistona teoreettiselle argumentaatiolle. Aineistopohjaisessa (engl. *corpus-based* tai *corpus-oriented*) tutkimuksessa sen sijaan kerätään kielitajun ohjaamana vartavastisesti laaja aineisto, jota analysoidaan ja josta saadaan jostain ilmiöstä esimerkiksi määrällistä tietoa ja teoriaa puoltavia tai kyseenalaistavia tietoja. Aineistovetoisessa (engl. *corpus-driven*) tutkimuksessa taas kielitajulla on pienin ennakkorooli. Siinä pyritään siihen, että laajasta aineistokokonaisuudesta itsestään nousisi teoreettisia ja hypoteettisia kysymyksiä tutkittavaksi (aineistovetoisuuden kritiikistä ks. esim. McEnery, Xiao & Tono 2006, 8–11). (Ks. tarkemmin Jantunen 2009.)

Metodit kehittyvät, ja niitä muun muassa yhdistellään keskenään. Esimerkiksi erilaisia kenttämenetelmiä (elisitaatiotekniikoita) voi

varioida monin tavoin tutkijan tavoitteiden ja strategian mukaan (esim. Cornips & Poletto 2005). Tyypillisen lomakepohjaisen tai suoran suullisen täsmäkyselyn lisäksi tutkittavan kielen puhujia voi muun muassa ohjata erilaisiin vuorovaikutuksellisiin tehtäviin, joissa puhujien huomaamatta todennäköisesti esiintyy tutkijan kaipaamia kielen rakenteita tai kielellistä toimintaa niiden aidossa käyttöympäristössä (ks. Kielten dokumentointi ja kieliopin kuvaus tk.). Joissain tutkimussuuntauksissa on myös luonnollista käyttää erilaisia aineistoja ja eri menetelmiä saman tutkimuskysymyksen avaamiseen (triangulaatio) (ks. esim. Plowright 2011; Creswell 2014; ks. Lingvistinen etnografia toimintayhteisöissä tk.).

Tutkimus jaetaan vanhastaan myös analyysitavan mukaan määrälliseen ja laadulliseen (ks. myös Johdannon luku 1 tk.). Määrällisessä tutkimuksessa ilmiötä pyritään usein kattamaan "ylhäältä käsin" ja näin hahmottamaan vallitsevaa tilaa kieli-, puhe- tai toimintayhteisössä (ks. Määrällinen korpuslingvistiikka; Kysely tk., tilastomenetelmistä kielentutkimuksessa muuten ks. esim. Luojola 2006; Pitkänen & Kohonen 1984). Laadullisessa tutkimuksessa lähdetään monesti "alhaalta käsin" tapauksista ja tilanteista ja pyritään hahmottamaan sitä, miten itse toiminnassa syntyy, rakentuu ja heijastuu yhteisö (ks. Laadullinen aineistopohjainen kielentutkimus; Lingvistinen etnografia toimintayhteisöissä tk.).[8] Esimerkiksi Töttö (2004, 9–17) kuitenkin luopuu koko eronteosta määrälliseen ja laadulliseen ja luonnehtii (sosiaali)tutkimuksen yleisjakoa termeillä *teoreettinen* ja *empiirinen*.

Kuvio 20 on sovellus ja mukaelma Tötön (2004, 10) kuviosta, joka hieman muokattuna sopii kuvaamaan myös kielentutkimusta. Sen ajatuksena on, että tutkimus ei useinkaan jakaudu tarkkarajaisesti teoreettiseen ja empiiriseen vaan niitä koskevaa metodiikkaa voi soveltaa yhdessä ja samassa tutkimuksessa tai sitten toinen niistä johtaa toiseen.




Kuvio 20. Kielentutkimuksen jakautuminen (soveltaen Töttöä 2004, 10).

Yleisesti ottaen tutkimuksenteko on enemmänkin kulkemista kehittyvän kysymyksenasettelun johdattelemana jollakin polulla tai kuvion 20 ilmentämällä kehällä, jolla ei välttämättä ole periaatteellistakaan päätepistettä. Jos hyväksytään kielentutkimuksen monimenetelmäisyys alana ja ajatus, että kaikkia metodeja tarvitaan, niin herää aiheellinen kysymys, ovatko tiukat metodiset erottelut aina tarpeellisia. Lisäävätkö ne aina ymmärrystä metodeista ja metodologiasta tavalla, joka auttaa kehittämään niistä entistä parempia, vai aiheuttavatko luokittelut pikemmin tiukoista koulukuntaeronteoista koituvia haittoja?

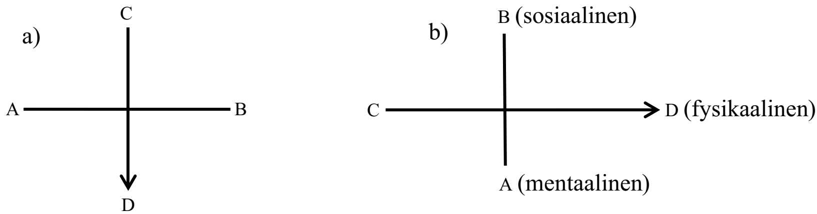
### 3.3.3. KIELI, AIKA, PAIKKA JA IHMINEN

Kieli ja kielentutkimuksen jonkin suuntauksen siitä lohkoma tutkimuskohde ovat aina erottamattomassa suhteessa aikaan. Kieli on olemassa ajassa, ja siksi kielentutkimuksen tulee pohtia, miten käsitteellistää aikaa ja kielen suhdetta siihen.

Klassisen jaon kielentutkimuksen suhteesta aikaan on esittänyt Ferdinand de Saussure (2014 [1916], 170–195) jaollaan staattiseen ja evolutiiviseen tieteeseen. Tätä kuvaa kuvio 21a. Siinä akseli ab kuvaa samanaikaisuutta. Se koskee sellaisia asioita ja ilmiöitä, jotka ovat olemassa yhtäaikaisesti tai samassa ajassa. Aikaa ei ole kuitenkaan tällä akselilla täsmällisesti määritetty. Kielentutkimuksessa tätä on vastannut synkroninen tutkimus. Vastaavasti akseli cd kuvaa peräkkäisyyttä, ja sillä

voidaan tarkastella jotakin akselille ab sijoittuvaa ilmiötä eri fysikaalisina hetkinä tai eri synkronisissa systeemeissä. Tätä on vastannut kielentutkimuksessa diakroninen tutkimus. Käsitteellisen eronteon synkroniseen ja diakroniseen kielentutkimukseen esittikin eksplisiittisesti ensimmäistä kertaa juuri de Saussure (2014 [1916]).

Yksinkertaistaen synkronisessa kielentutkimuksessa tutkitaan kielen yhtenäiseksi hahmotettua tilaa jollakin ajan hetkellä. Tämä tila voi sijoittua johonkin historialliseen aikaan (esim. myöhäiskantasuomi) tai nykyhetkeen (esim. suomen kirjoitettu yleiskieli). Aika käsitetään tällöin yleensä abstrahoituna ja ikään kuin pysäytettynä ja etäännytettynä. Kielen tila on eräänlainen kohtisuora leikkaus fysikaalisesta aikajanasta, ja sen ajatellaan koskevan yleensä kieltä kokonaisuudessaan järjestelmänä. Vastaavasti diakroninen kielentutkimus keskittyy kielen muutokseen ajan funktiona. Tämä tapahtuu yleensä vertailemalla kielen tiloja eri synkronisina hetkinä. (Ks. Historiallinen ja vertaileva kielentutkimus tk.)




Kuvio 21. (a) Tieteiden akselit de Saussuren (2014 [1916], 171) mukaan; (b) konventionaalinen aikajana (cd = fysikaalinen taso) ja synkroninen leikkaus (a = mentaalinen taso, b = sosiaalinen taso) (vrt. kuvio 15).

Vaikka tällaisissa kielitieteen aloja selittävissä malleissa käsite *aika* toimii heuristisesti erilaisten kielentutkimuksen lähestymistapojen eroja selvennettäessä, ei tule kuitenkaan unohtaa, että aika itsessään ei ole ontologisesti mitenkään ongelmaton käsite. Kantilaisessa teoriassa aika muodostaa ihmisen tietoisuuden ja havainnon ennakkoehdon ja rakenteen: ihminen voi havainnoida ja ajatella vain ajassa.

Arkisen aika-käsitteen kautta voidaan kuitenkin kuvata kielellistä tapahtumista ja muuttumista. Kuviossa 21b, jota de Saussure ei

sellaisenaan itse esitä, akselisto kuviossa 21a on käännetty niin, että akseli cd on horisontaalinen ja vastaa (usein länsimaisessa ajattelussa) konventionaalisesti esitettyä aikajanaa tai -nuolta kuvaamassa fysikaalista aikaa. Akseliin cd kiinnittyvät fysikaalisessa ajassa aikapaikkaisesti ja kollektiivisesti ilmenevät ja kehittyvät tapahtumat. Välittömästä akselista cd irtautuvat asiat ovat tavalla tai toisella abstraktioita. Abstrahoinnin suunta cd:stä kohti b:tä kuvaa sosiaalista ja suunta kohti a:ta mentaalista (vrt. kuvio 15).[9] Myös nämä ovat toki ajallisia, mutta niitä ei voi suoraan ja kollektiivisesti havainnoida ajallis-paikallisesti kuten akselille cd sijoittuvia ilmiöitä. Mentaalinen ilmenee suoraan vain tietoiselle yksilölle itselleen, kun taas muille se ilmenee epäsuorasti. Myös sosiaalinen normien tai konventioiden maailma ilmenee vain epäsuorasti. Havainnot koskevat siis primaaristi vain akselia cd. Akselien leikkauspiste on käsitteellinen nyt-hetki, jolloin suunta c:hen on historiaa ja suunta d:hen tulevaisuutta. Seuraava esitys kytkeytyy kuvioon 21b.

Luonnollisesti selvin suhde aikaan on kaikella kielen muuttumista ja tämän muutoksen yleisiä periaatteita kuvaavalla historiallisella tutkimuksella. Tällöin ajallinen ulottuvuus on usein kymmeniä, satoja tai jopa tuhansia vuosia. Aikavälejä voidaan periodisoida ja kieliä tai kielen esi- ja kantamuotoja nimetä (esim. varhais-uusenglanti noin 1500–1710, indoeurooppalainen kantakieli yms.). Kielten kirjoitetun historian aikana on mahdollista pysytellä c:llä absoluuttisissa ajoissa tai arvioissa niistä ja tutkia esimerkiksi jonkin kieliopillisen ilmiön reaalista ilmentymistä ja kehittymistä ajan saatossa säilyneiden tekstien perusteella (ks. Näkökulmia filologiseen tutkimukseen tk.). Tämä mahdollistaa myös holistisemman historiallisen sosiolingvistisen tarkastelun. Koska kielen muutos on yleensä aina vähittäistä, sen hahmottaminen synkronisista ilmenemisistä on kuitenkin enemmän tai vähemmän teoreettista abstrahointia. (Ks. Historiallinen korpuslingvistiikka tk.; Nyman 2000.) Tällainen tutkimus voi pysytellä yhden kielen kehityskaaren sisällä (esim. muinaisenglannista nykyenglantiin) tai kuvata kehitystä äitikielestä sisarkieleen (esim. vulgaarilatinasta italiaan). Tutkimuksen toteutettavuuteen vaikuttaa se, kuinka pitkältä aikaväliltä, kuinka tiheästi, kuinka paljon ja millaisista tekstilajeista tekstiaineistoja on käytettävissä. Tietotekninen korpustyö ja tilastolaskenta toimivat luonnollisina apuina.

Kielten kirjoittamaton historia, jota historiallis-vertaileva tutkimus tarkastelee, pyrkii kurottamaan yleensä kielen varhaishistoriaan. Se voi olla historiaa, josta reaalista kirjallista todistusaineistoa on hyvin niukasti (esim. tuohikirjoituksia tai muita kielen muistomerkkejä) tai ei ollenkaan, tai tunnetun historian takaista esihistoriaa ennen kirjoitusjärjestelmiä ja kirjoitettua kieltä. Kielen esihistoriaa pyritään kartoittamaan pääasiassa vertailemalla eri synkronisia kieliä toisiinsa ja muodostamalla tämän perusteella rekonstruktioita. Tällöin selvitellään muun muassa kielten sukulaisuussuhteita, kielikuntaluokitteluja, lainasanalähteitä ja -kerrostumia ja yksittäisten kielten kehittymistä ja vaiheita sekä hahmotellaan kantakieliä ja erilaisia ajoituksia (ks. Historiallinen ja vertaileva kielentutkimus tk.). Tällainen tutkimus (esim. äännehistoria) kuvaa esimerkiksi sukukielivertailun perusteella teoretisoituja äänteiden ja äännejärjestelmien rekonstruktioita, joita voidaan suhteuttaa absoluuttisiin aikoihin usein vain muiden reaalitieteiden (esim. arkeologia, kansatiede, migraatiotutkimukset) avulla. Muuten kyseessä on esimerkiksi kantakielten ajoituksia ajatellen ajallinen relatiivisuus, joka suhteuttaa eri abstrahoituja rekonstruktiotasoja (eli kielen synkronisia leikkauksia) kronologiseen järjestykseen. Lisäksi varhaiset rekonstruktiot, esimerkiksi perimmäisimmät kantakielet, ovat vieläpä rekonstruktioiden rekonstruktioita eli viime kädessä hyvin teoreettisia käsitteitä.

Myös sisäinen rekonstruktio paljastaa teoretisoidun synkronisen kielijärjestelmän sisällä relatiivisesti kronologisia ilmiöitä. Se kuvaa tavallaan historiallista näennäisaikaa synkronisen systeemin sisällä (reaali- ja näennäisajasta ks. Sosiolingvistiikka tk.). Sekä historiallis-vertaileva metodi että sisäinen rekonstruktio edustavat teoretisointia, joka etäännyttää c:stä. Usein kyse on yksittäisten ilmiöiden vertailusta kahden tai useamman abstraktin synkronisen tason kesken. Erikseen on vielä mainittava nimistön asema keskeisenä kielen historiaa säilyttävänä osa-alueena (ks. Nimistöntutkimus tk.).

Lähemmäs nykyhetkeä eli akselien leikkauskohtaa sijoittuva tai nykyhetkeen suhteutettu kielentutkimus voi myös olla perspektiiviltään historiallista. Tällöin kuitenkin aikajänteet ja periodit ovat huomattavasti lyhyempiä (esim. joitain kymmeniä vuosia) ja yleensä vertaillaan yhtäaikaisesti elossa olevien sukupolvien kieltä. Kieli myös kiinnitetään ajan

ja paikan ohella yhä kiinteämmin eri tavoin määrittyviin ihmisryhmiin.

Sosiolingvistiikan reaaliaikatutkimukset kiinnittyvät absoluuttiseen aikaan noin 10–20 vuoden ajalta tai tiheämmin. Paneelitutkimuksessa samoja yksilöitä tutkitaan eri aikoina (esim. 10 vuoden välein). Trenditutkimuksessa tutkittavat muuttujat säilyvät mutta tutkittavat henkilöt vaihtuvat eri aikoina (esim. 10 vuoden välein). Koska perinteinen piirrevariaatioon keskittyvä sosiolingvistiikka tarkastelee kielenulkoisin muuttujin (esim. ikä, sukupuoli, asuinpaikka) määrittyvien ihmisryhmien kieltä, se abstrahoi kuvauksen sosiaaliseen tasoon (b). Yksilön idiolektin muuttumisen tai muuttumattomuuden tutkimus mahdollistaa myös tapauskohtaisesti yksilön kieltä ja jopa mentaalikielioppia koskevien oletusten tekemisen (a). Näennäisaikatutkimus, jossa ennustetaan seuraavan sukupolven kielen muuttumista samaan aikaan elossa olevien edellisten sukupolvien kielen perusteella, tarjoaa hypoteeseja jo tulevaisuuden kielestä. (ks. Sosiolingvistiikka; Haastattelu aineistonkeruun metodina sosiolingvistiikassa tk.)

Tulevaisuuskin (d) on kielentutkimukselle relevanttia. Teoreettisessa kielentutkimuksessa tämä näkyy esimerkiksi kieliopillistumisen eli grammatikalisaation tutkimuksessa. Siinä esitetään kieliopilliset muutosprosessit peräkkäisinä osatapahtumina ja usein jatkumomaisena. Niistä esitettyjä teoreettisia yleistyksiä voi pitää hypoteettisina ennusteina vastaavanlaisista muutoksista tulevaisuudessa. Tähän liittyy myös kielten universaalien tutkimus, sillä esitettyjen kieltä koskevien yleistysten ajatellaan vaikuttavan myös tulevaisuudessa puhuttavissa kielissä. Tällaisina kieliopillistumisen ja kielen typologisten universaalien voi ajatella edustavan ennustavan selittämisen malleja tai hypoteeseja kielentutkimuksessa.

Kuitenkin konkreettisemmin ja pysäyttävämmin tulevaisuus näkyy kielisosiologiassa esimerkiksi kielten katoamisessa. On arvioitu, että kuluvan satavuotisjakson aikana vähintään puolet ja jopa 90 % maailman arviolta noin 6 000–7 000 puhuttavasta kielestä lakkaisi olemasta (Bradley & Bradley 2002). Kun jotain kieltä puhuvat yksilöt kuolevat (a) eikä geneettistä ja kielellistä uusiutumista tapahdu riittävästi, kielen sosiaalinen asema on heikko eikä tilannetta pystytä tai ehditä kääntämään paremmaksi (b). Tällöin kokonainen kieli saattaa lakata olemasta. Usein kyseeseen tulee sekin, että puhujien kielitaito alkaa rappeutua ja päädytään kielenvaihtoon. Tällaisen tapahtuman myötä menetetään aina

kunkin kielen kohdalla periaatteessa mittaamaton määrä tietoa ja ihmisen kulttuuriperimää. Siksi kielellis-kulttuurisen monimuotoisuuden korostamisen tulisi nousta yhdeksi kielentutkimuksen globaaleista poliittisista tavoitteista. (ks. Antropologinen lingvistiikka; Kielikontaktien tutkimus; Kielten dokumentointi ja kieliopin kuvaus tk.)

Paikkaa tai kielialueen sijaintia voi pitää ajan rinnalla toissijaisena kieltä määrittävänä tekijänä erityisesti historiallis-vertailevassa suuntauksessa, sillä kielet ja kieliyhteisöt voivat aikojen saatossa alueellisesti laajentua, supistua tai vaihtaa maantieteellistä paikkaansa. Sen sijaan esimerkiksi kielikontaktien tutkimus lokalisoi kieliä yhtä lailla ajallisesti kuin paikallisesti (ks. Kielikontaktien tutkimus tk.). Paikallisesti lähekkäin ja luonnollisessa yhteydessä olevien erikielisten kieliyhteisöjen kontaktit synnyttävät esimerkiksi sellaisia ilmiöitä kuin kielijatkumot (esim. Pohjois-Afrikan, Arabian niemimaan ja Lähi-idän arabian murteet), kieliliitot (esim. Balkan), pidgin- ja kreolikielet ja lingua francat sekä pysyvät kielenvaihdot ja tilanteiset koodinvaihdot. Paikkaan viittaa myös termi *areaaliset piirteet*. Samoin ekolingvistiikka tutkii sitä, miten muun muassa kieliyhteisön fyysinen ympäristö näkyy kielessä (ks. Antropologinen lingvistiikka tk.). Kontaktien historiallinen aspekti näkyy synkronisesti kielissä esimerkiksi lainasanakerrostumina (substraatit). On kuitenkin hyvä muistaa, että paikka ei sekään ole kuitenkaan itsestään selvä ontologinen kategoria vaan käsitteellinen konstruktio, jonka jokaisen tutkijan on omassa tutkimuksessaan määriteltävä aina erikseen.

Kielen systeeminen luonne ja sen erilainen (paikallinen, tilanteinen, roolittainen, ihmisryhmittäinen jne.) luonnollinen variaatio (ks. kuvio 16) sisältää potentiaalisia muutoksen siemeniä, joten historiallinen perspektiivi on koko ajan piilevä osa kielen elämää. Se pakottaa määrittelemään esimerkiksi juuri variaation taustakäsitteitä uudelleen kielisysteemin ulkoisin tekijöin (esimerkiksi *diskurssirooli*, *identiteetti*, *indeksisyys*). Näin toimivaa tutkimusta edustavat esimerkiksi keskustelunanalyysi, tekstintutkimus, diskurssintutkimus, lingvistinen etnografia ja antropologinen kielentutkimus (ks. ko. lukuja tk.). Lisäksi osassa kielentutkimusta merkitys nähdään ennen kaikkea semioosina eli merkitsemistapahtumana koodin sijaan ja usein osana kulttuurisia ja ideologisia taustoja, jotka ovat historiallisia. Tutkijan on tällöinkin

mielekästä tarkistaa käyttämänsä taustateorian ontologisia oletuksia. Esimerkiksi funktionaalisella sosiosemiotiikalla (esim. Halliday 1978), kognitiivisella semiotiikalla (esim. Zlatev 2012; Sinha 2014) ja evolutiivisella semiotiikalla (esim. Cariani 1998) on erilaiset käsitykset todellisuuden luonteesta. Kun lisäksi de Saussuren kaksipaikkaisesta merkistä siirrytään monipaikkaisiin merkkeihin (esim. Peirce, Derrida, Kristeva), siirrytään usein myös noiden merkkien merkityksiin vaikuttaviin – usein dialektisiin – prosesseihin, joita ei voi jättää tulkinnan ulkopuolelle.

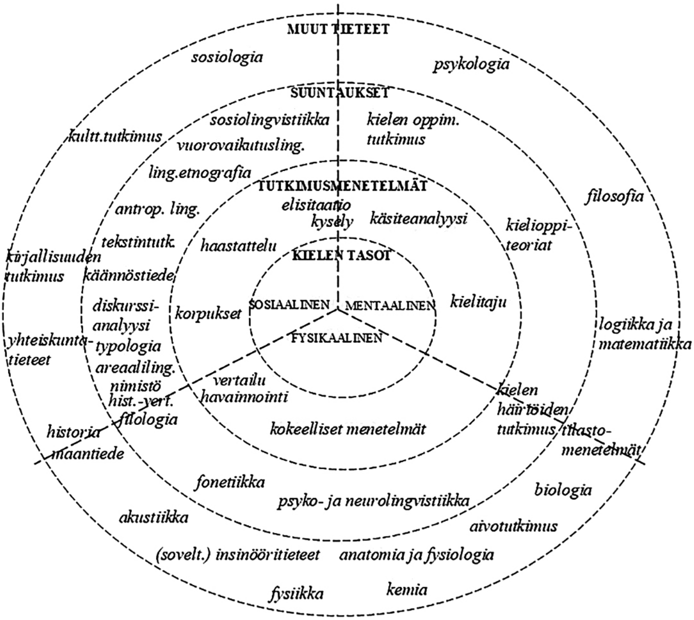
Lukuun ottamatta ehkä yksilökielen eli idiolektin tutkimusta ajan, paikan ja ihmisen teoretisointi on koskenut valtaosin sosiaalista tasoa (b). Sellainen reaaliaikainen tutkimus, jossa pyritään kokeellisesti selvittämään esimerkiksi puhujan kielen tuotto- tai kuulemiskykyä, kognitiivista hahmottamista tai aivojen neurologista toimintaa, pysyttelee usein käytännön syistä hyvin pienen aikaikkunan sisällä (ks. Kokeellinen fonetiikka; Psyko- ja neurolingvistiikka tk.). Tutkimusaineisto on tällöin joistain sekunneista millisekunteihin. Saatavan aineiston varassa on mahdollista esittää malleja kieltä koskevan aivotoiminnan edustumisesta (cd) sekä viime kädessä arvioida teoretisointeja mentaalikielioppimallien (a) psykolingvistisesta realistisuudesta. Erityisesti perinteiset ja uudemmat kielioppikuvaukset ja kielenkuvausmallit ovat tavallisesti myös ikään kuin ajattomia (atemporaalisia). Ne kuvaavat koko kielioppijärjestelmää tai jotakin sen osaa usein joko sosiaalisesti vallitsevana (b) tai sellaisena kuin sen oletetaan tai teoretisoiden esitetään edustuvan kielenpuhujien mielessä (a) tämän omaksuttua sosiaaliset normit (ks. Kieliopin tutkimus ja kielioppiteoriat tk.).

## 3.4. Kielen tasot ja tutkimuksen monimenetelmäisyys

Kielentutkijan on tutkimuksessaan jotenkin määriteltävä luonnollinen kieli, tutkimuksensa kohde. Tässä tehtävässä on mielekästä muistaa, että ihmiselle ilmenevä ja ihmisen käsitteistämä todellisuus sekä tutkimuksessa käytetyt muuttujat, kategoriat, rakenteet ja muodot eivät ole välttämättä sama asia kuin ontologia. Luvussa 3.2.3. käsiteltiin tutkimuskohteen ontologisointia: tutkijan taustateoria vaikuttaa siihen, millaisen eksistenssin hän osoittaa tutkittavalle ilmiölle.

Tutkimuksessa on luonnollista käsitteellistää osa kulloinkin tarkasteltavasta maailmasta ja esittää erilaisia painotuksia sille, mitä puolta kielestä halutaan ensi sijassa tarkastella: kielen konventionaalistunutta koodin rakennetta tai sisältöä (eli symbolisen yksikön muotoa tai merkitystä), kieltä kokonaisena abstraktina järjestelmänä tai kielen rakenteiden esiintymiä käytössä ja näiden funktioita ja tulkintoja vai kieltä tietona tai käyttäytymisenä osana muuta inhimillistä toimintaa. Kieltä voidaan myös suhteuttaa käsiteistettyyn aikaan ja paikkaan sekä sitä käyttävään ihmiseen eri tavoin (ks. Johdannon luku 3.3.3. tk.).

Kuviossa 22 esitetään karkea kuva kielentutkimuksen menetelmien tai suuntausten ja kielen sekä eräiden muiden tieteiden suhteista. Kuvio ei ole kattava, ja sitä on syytä tulkita heuristisena mallina, jolla voidaan yksinkertaistaa monimutkaista ja monitasoista suhteiden järjestelmää. Kuvio muodostuu sisäkkäisistä ympyrän kehistä, joiden rajat eivät ole jyrkät. Keskiössä on kieli primaarina tutkimuskohteena ja sen jakautuminen fysikaaliseen, mentaaliseen ja sosiaaliseen tasoon (vrt. kuviot 15 ja 21b).[10]




Kuvio 22. Kielen tasot, tutkimusmenetelmät ja kielentutkimuksen suuntaukset.

Kielitieteen menetelmiä ja kielentutkimuksen suuntauksia ei voi useinkaan selvästi ja yksioikoisesti erottaa toisistaan. Suuntauksissa kiteytyy usein jonkin menetelmän perinteinen tietämys ja tällä metodilla saadusta tiedosta jalostunut kielikäsitys ja tieteellinen teoria. Jotkin metodit myös käytännössä hahmotetaan samalla suuntauksiksi (esim. historiallis-vertaileva metodi tai suuntaus). Suuntausten tietynlainen toissijaisuus koskee niiden tieteensosiologista jäsentymistä, paradigmatisoitumista, erilaisia ohjausvaikutuksia (esim. mikä on kulloinkin arvostettavana pidettävää ja tavoiteltavaa tutkimusta, prominenttien tutkijoiden valtaa jne.) ja esimerkiksi yksittäisten tutkijoiden ja opiskelijoiden asemaa ja sosiaalistamisprosessia tiedeyhteisössä. Tällaiset seikat ovat enemmän tiedeyhteisöä jäsentäviä tekijöitä ja viime kädessä toissijaisia tekijöitä tutkimuksen ytimen (tiedonkohde, metodi ja metodologia) rinnalla, vaikka niillä onkin merkittävää vaikutusta myös tehtävään tutkimukseen (esim. tutkimustrendien tai pitkäaikaisten tutkimusperinteiden muodossa). (Ks. esim. Karlsson 1975; Kelomäki 2009.) Tieteensisäisen järjestäytymisen lisäksi tutkimuksen muotoutumiseen vaikuttavat toki myös tieteen ulkoiset, yhteiskunnalliset tekijät ja esimerkiksi kielipolitiikka (esim. Lappalainen, Sorjonen & Vilkuna 2010). Tähän liittyy se, kuinka kieli yhteiskunnassa yleisesti ymmärretään, millainen merkitys sille annetaan, mitä tutkimukselta toivotaan ja millaisen tutkimuksen katsotaan kulloinkin olevan tässä suhteessa tuottavinta. Nämä seikat voivat vaikuttaa myös paikallisesti painotusten siirtymiin tieteenalan sisällä.

Tutkijan on mielekästä analysoida oman metodinsa ja taustateoriansa käsityksiä kielestä (esim. kieli järjestelmänä, kieli toimintana yms.). Ilmiöihin pureutuvat kysymyksenasettelut ohjaavat metodivalintaa, minkä seurauksena kieleen kohdistuu jokin näkökulma, jonka voi pelkistää tutkimuskohtaisesti joksikin kielikäsitykseksi. Eri suuntausten tai metodien piirissä voi olla esimerkillisiä mallitutkimuksia eli eksemplaareja, joita voidaan seurata tutkimuksenteossa. Tällainen oli aikanaan esimerkiksi E. N. Setälän vuonna 1891 ilmestynyt teos *Yhteissuomalainen äännehistoria*, joka toimi pitkään fennougrististen äännehistoriakuvausten malliesityksenä ja joka on monessa suhteessa pätevä yhäkin. Yleisesti mallitutkimuksissa on aina vaarana se, että ne sitovat tutkimusalaa liikaa tai niiden sisältämiä virheitä toistetaan uusissa tutkimuksissa.

Kielitieteellisten teorioiden sijoittaminen ontologisiin näkemyksiin on toisaalta hankalaa siksi, että kielitieteen metateoreetikot eivät useinkaan ole määritelleet omia ontologisia käsityksiään selvästi suhteessa käytyihin koulukuntakeskusteluihin. Toisaalta näissä määritelmissä on usein sellaisia ristiriitaisuuksia, jotka tekevät ko. teoriasta jollain tavoin sisäisesti ristiriitaisen. Esimerkiksi Langacker (1987, 20, 157, 178) käyttää termejä *holismi* ja *emergenssi* mutta toisaalta olettaa, että kognitiiviset ilmiöt voidaan redusoida eli palauttaa neurokemialliseksi toiminnaksi (emergenssistä ks. Määttä, Nieminen & Pälli 2000). Hän tekee siten kolme keskenään ristiriitaista sitoumusta (ks. Kenttä 2003, 78). Edes toisiaan hyvin lähellä olevien teorioiden, kuten prototyyppiteorian (Rosch 1978; 1999) ja kognitiivisen semantiikan (Lakoff & Johnson 1999), ontologiset oletukset eivät ole tarkalleen ottaen samat.

## 3.5. Lopuksi

Olemme tässä luvussa kehittäneet yleiskuvaa kielentutkimuksen kentästä ja kiinnittäneet huomiota joihinkin metateoreettisen tarkastelun kannalta keskeisiin käsitteellisiin lähtökohtiin. Tieteenalan monimenetelmäisyys perustuu siihen tosiasiaan, että kielen kaltaista hyvin mutkikasta ilmiötä voidaan lähestyä useasta eri tarkastelukulmasta ja se voidaan hahmottaa jo perusolemukseltaankin moni-ilmeiseksi (esim. kielen fysikaalinen, mentaalinen ja sosiaalinen taso, ks. kuviot 15, 21b ja 22). Tällaisen selvittely edellyttää välttämättä usean metodin käyttöä. Huomattava on kuitenkin ensinnä se, että kielen käsitteellistäminen tutkimustyössä on sikäli aina teoriasidonnaista, että vasta taustateoria määrittelee ontologisia sitoumuksia. Toiseksi kielen ontologisointi eli kielen olemuksen näkeminen ensisijaisesti yhdenlaisena ei millään välttämättömyydellä johda normatiivisesti tietynlaiseen metodologiaan. Kielentutkimuksen käytäntö ja sitä kuvaava tieteenfilosofia ovat toistaiseksi pysytelleet melko etäällä toisistaan (ks. kuitenkin esim. Hamunen, Nieminen, Kelomäki & Dufva 2019). Hyvin toivottavaa olisi, että tulevaisuudessa kielentutkimuksen metateoriaa tutkittaisiin enemmän.

**Aiheesta lisää:**

Scholz, Barbara C., Pelletier, Francis Jeffry & Pullum, Geoffrey K. 2016. Philosophy of linguistics. Julkaisussa: Zalta, Edward N. (toim.) *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). [verkkoaineisto]. [viitattu 25.5.2018]. Saatavissa: https://plato.stanford.edu/archives/win2016/entries/linguistics/.

Stainton, Robert J. 2018. *Philosophy of Linguistics*. Oxford Handbooks Online. [verkkoaineisto]. [viitattu 25.5.2018]. Saatavissa: http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935314.001.0001/oxfordhb-9780199935314-e-002.

## VIITTEET

1 Davidsonin käsitys perinteisestä tai vallitsevasta kielentutkimuksesta oli selvästi ennen kaikkea chomskylainen generativistinen näkemys kielestä järjestelmänä.

2 Niin kuin monien filosofien ajatukset kehittyvät ajan myötä, näin kävi myös Popperille. Hänen rationalistiseen ajatteluunsa sisältyi hänen myöhäisfilosofiassaan yhä enemmän realistisia elementtejä. Popper ei itse käsitellyt tätä dilemmaa omassa ajattelussaan, joten tarkkaa selkoa ovat tehneet vasta Popper-tulkitsijat (ks. esim. John Wettersten (1992) *The Roots of Critical Rationalism*).

3 On hyvä muistaa, että de Saussuren *Yleisen kielitieteen kurssi* on postuumi julkaisu opiskelijoiden muistiinpanojen pohjalta (ks. Nyman & Suominen 2002). Edellä esitetyn englanninkielisen sitaatin suomenkielinen vastine kuuluu näin: ”Merkit, joista kieli koostuu, eivät ole abstraktioita, vaan jotakin todella olemassa olevaa. Juuri merkkejä ja niiden suhteita kielitiede tutkii. Voimme kutsua niitä tämän tieteen *konkreettisiksi olioiksi*.” (de Saussure 2014 [1916]: 201.) Toisaalla de Saussure luonnehtii merkkiä tähän tapaan: ”Kieli [langue] on luonteeltaan yhtä konkreettinen asia kuin puhuntakin [parole], ja tästä on suurta etua tutkimuksessa. Vaikka kielelliset merkit ovatkin olennaisesti psyykkisiä [!], ne eivät ole abstraktioita.” (Mts. 89.)

4 *Lektin* käsitteellä (esim. *idiolekti*, *sosiolekti*, *etnolekti*) on sosiolingvistiikassa tarkoitettu erilaisia kielen varieteetteja, jotka ovat hahmotettavissa ilmiasultaan tietynlaisiksi kielen muodoiksi (esim. aluemurre).

5 *Elisitointi* (*elisitaatio*, *kirvoittaminen*) on kielentutkimuksessa yhteisnimitys sellaisille menetelmille, joissa kielenpuhujat pyritään saamaan tuottamaan tutkijan toivomia kielen rakenteita tai ilmauksia tai ylipäätään toivottua kielellistä toimintaa joko tietoisesti tai tiedostamattomasti.

6 *Solipsismilla* viitataan siihen, että vain yksilön tietoinen tajunta ja mielensisällöt ovat ainoita täysin varmoiksi tiedettyjä asioita.

7 *Ekologinen validiteetti* tarkoittaa sitä, että tutkimuksen pääargumentti populaation suhteen on oikeutettu aineiston ja metodin kautta. Jos esimerkiksi on tutkittu ylioppilaskirjoituksia, ei tuloksia voida yleistää kaikkiin tekstilajeihin ja ikäluokkiin.

8 Jako *mikro-* ja *makrotutkimukseen* kuvaa tätä eroa jollain tavoin siedettävästi, mutta sekin on dikotomia, jonka voi helposti tulkita liian yksioikoisesti.

9 Kuvio 21b yhdistää oikeastaan kaksi asiaa: de Saussuren esittämän kuvion (ks. kuvio 21a) käännetty akselisto yhdistetään kuviossa 15 esitettyjen kielen heurististen olemustasojen kanssa.

10 Tasojen sijaan voisi ehkä puhua *kielen olomuodoista*, jolloin terminologinen ero tulisi selvemmäksi: *kielen olomuodot* (fysikaalinen, mentaalinen, sosiaalinen), *kielen* (*kuvaus-*) *tasot* (perinteinen jako fonologiaan, morfologiaan ym.) ja *kielen kanavat* (ilmeneminen kirjoitettuna, puhuttuna, viitottuna). Epäselvyyden vaaraa ei tässä yhteydessä kuitenkaan ole.

## KIRJALLISUUS


Arppe, Antti & Järvikivi, Juhani. 2007. Take empiricism seriously! In support of methodological diversity in linguistics. *Corpus Linguistics and Linguistic Theory* 3:1, 99–109.

Baker, Gordon P. & Hacker, Peter Michael Stephan. 1988. *An Analytical Commentary on the Philosophical Investigations. Volume 2. Wittgenstein: Rules, Grammar and Necessity*. Oxford: John Wiley & Sons.

Bloomfield, Leonard. 1957 [1933]. *Language*. London: George Allen & Unwin Ltd.

Bradley, David & Bradley, Maya. 2002. Introduction. Julkaisussa: Bradley, David & Bradley, Maya (toim.) *Language Endangerment and Language Maintenance: an Active Approach*. London: Curzon Press, xi–xx.

Brattico, Pauli. 2008. *Biolingvistiikka: johdatus kielen biologiaan, evoluutioon ja kognitiivisiin perusteisiin*. Helsinki: Gaudeamus.

Cariani, Peter. 1998. Towards an evolutionary semiotics: the emergence of new signfunctions in organisms and devices. Julkaisussa: Vijver, Gertrudis Van de, Salthe, Stanley N. & Delpos, Manuela (toim.) *Evolutionary Systems. Biological and Epistemological Perspectives on Selection and Self-Organization*. Dordrecht: Springer, 359–376.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. New York (NY): Praeger.

Chomsky, Noam. 1989. *Language and Mind*. The Darwin Lecture, Darwin College.

Chomsky, Noam. 2001 [1965]. *Aspects of the Theory of Syntax*. (21. p.). Cambridge (MA): MIT Press.

Chomsky, Noam. 2007. *On Language. Chomsky's Classic Works Language and Responsibility and Reflections on Language*. New York (NY): The New Press.

Cornips, Leonie & Poletto, Cecilia. 2005. On standardizing syntactic elicitation techniques. *Lingua* 115, 939–957.

Creswell, John. 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. (4. p.). Los Angeles (CA): Sage.

Crystal, David. 2008. *A dictionary of linguistics and phonetics*. (2. p.). Malden (MA): Blackwell.

Davidson, Donald. 2006 [1986]. A nice derangement of epitaphs. Julkaisussa: Davidson, Donald (toim.) *The Essential Davidson*. Oxford & New York (NY): Clarendon Press/Oxford University Press, 251–265.

Dennett, Daniel C. 1999. *Tietoisuuden selitys*. [*Consciousness explained*]. Suomentanut Tiina Kartano. Jyväskylä: Art House.

Fagan, Stephen & Gençay, Ramazan. 2010. An introduction to textual econometrics. Julkaisussa: Ullah, Aman & Giles, David E. A. (toim.) *Handbook of Empirical Economics and Finance*. Boca Raton: Chapman & Hall/CRC Press, 133–153.

Grice, Paul H. 1975. Logic and conversation. Julkaisussa: Cole, Peter & Morgan, Jerry L. (toim.) *Syntax and Semantics 3. Speech Acts*. New York (NY): Academic Press, 41–58.

Haddington, Pentti & Sivonen, Jari (toim.). 2010. *Kielentutkimuksen modernit klassikot. Kognitiivinen ja funktionaalinen kielitiede*. Helsinki: Gaudeamus.

Halliday, Michael A. K. 1978. *Language as a Social Semiotic: The Social Interpretation of Language and Meaning*. London: University Park Press.

Hamunen, Markus, Nieminen, Tommi, Kelomäki, Tapani & Dufva, Hannele (toim.). 2019. *Käänteitä ja käsitteitä. Näkökulmia kielitieteelliseen keskusteluun*. SKST 1454. Helsinki: SKS.

Harris, Roy. 1988. *Language, Saussure and Wittgenstein. How to Play Games with Words*. London: Routledge.

Itkonen, Esa. 1997. The social ontology of linguistic meaning. *Yearbook of the Linguistic Association of Finland* 10, 49–80.

Itkonen, Esa. 2019. Concerning the scope of normativity. Julkaisussa: Mäkilähde, Aleksi, Leppänen, Ville & Itkonen, Esa (toim.) *Normativity in Language and Linguistics*. Studies in Language Companion Series 209. Amsterdam: John Benjamins, 29–67.

Itkonen, Esa & Pajunen, Anneli. 2010. *Empiirisen kielitieteen metodologia*. Suomi 199. Helsinki: SKS.

Jantunen, Jarmo Harri. 2009. Ei pelkästään mielikuvituksen puutteen vuoksi – kieliaineistojen systemaattinen käyttö kielentutkimuksessa. *Virittäjä* 113:1, 101–113.

Kaila, Eino. 1939. *Inhimillinen tieto. Mitä se on ja mitä se ei ole*. Helsinki: Otava.

Karlsson, Fred. 1982. *Suomen kielen äänne- ja muotorakenne*. Porvoo: WSOY.

Karlsson, Fred. 1975. Fennistiikan tieteenparadigmasta ja sen ohjausvaikutuksesta. *Virittäjä* 79:2, 179–193.

Kauhanen, Irina. 2006. Norms and sociolinguistic description. A Man of Measure Festschrift in Honour of Fred Karlsson. *SKY Journal of Linguistics* 19, 34–46.

Kelomäki, Tapani. 2009. *Nuorgrammaattisuuden myytti suomen kielen tutkimuksessa*. Helsinki: Helsingin yliopiston suomen kielen ja kotimaisen kirjallisuuden laitos.

Kelomäki, Tapani. 2019. Kielitieteen käänteistä. Julkaisussa: Hamunen, Markus, Nieminen, Tommi, Kelomäki, Tapani & Dufva, Hannele (toim.) *Käänteitä ja käsitteitä. Näkökulmia kielitieteelliseen keskusteluun*. SKST 1454. Helsinki: SKS, 31–60.

Kenttä, Reetta-Leena. 2003. Langackerin kognitiivisen kieliopin metateoreettinen analyysi: *Kognitio*-käsite ja tieteellinen argumentaatio kognitiivisessa kielitieteessä. Pro gradu -tutkielma. Helsinki: Helsingin yliopisto.
Konstenius, Reetta. 2014. *Empiria, eksperimentti ja etiikka. Kielitieteen metateoriaa*. Väitöskirja. Helsinki: Helsingin yliopisto.
Lakoff, George. 1987. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. Chicago & London: The University of Chicago Press.
Lakoff, George & Johnson, Mark. 1999. *Philosophy in the Flesh*. New York (NY): Basic books.
Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Volume I. Theoretical Prerequisites*. Stanford (CA): Stanford University Press.
Lappalainen, Hanna, Sorjonen, Marja-Leena & Vilkuna, Maria (toim.). 2010. *Kielellä on merkitystä. Näkökulmia kielipolitiikkaan*. SKST 1262. Helsinki: SKS.
Luojola, Teemu. 2006. *Kielitieteellisen aineiston kvantitatiiviset analyysimenetelmät*. [verkkoaineisto]. [viitattu 1.2.2019]. Helsingin yliopisto, yleinen kielitiede. Saatavissa: http://www.ling.helsinki.fi/~fkarlsso/methods/kvant_men.pdf.
McEnery, Tony, Xiao, Richard & Tono, Yukio. 2006. *Corpus-based Language Studies: An Advanced Resource Book*. London: Routledge.
Mäkilähde, Aleksi & Hynönen, Emmi. 2019. Normatiivisuus ja normin käsite kognitiivisessa kieliopissa, peruskielioppiteoriassa ja systeemis-funktionaalisessa kieliopissa. Julkaisussa: Hamunen, Markus, Nieminen, Tommi, Kelomäki, Tapani & Dufva, Hannele (toim.) *Käänteitä ja käsitteitä. Näkökulmia kielitieteelliseen keskusteluun*. SKST 1454. Helsinki: SKS, 267–304.
Määttä, Urho. 1994. *Funktionaalinen selittäminen morfologiassa. Metateoriaa ja huomioita suomen ja sen sukukielten tutkimusperinteestä*. Opera Fennistica & Linguistica 7. Tampere: Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos.
Määttä, Urho. 2000a. Mistä on pienet säännöt tehty? *Virittäjä* 104:2, 203–221.
Määttä, Urho. 2000b. Emergentistinen kielitiede. Julkaisussa: Määttä, Urho, Nieminen, Tommi & Pälli, Pekka (toim.) *Emergenssin kielelliset kasvot*. Folia Fennistica & Linguistica 24. Tampere: Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos, 17–39.
Määttä, Urho. 2000c. Kielitieteen emergenttinen metateoria. *Virittäjä* 104:4, 498–517.
Määttä, Urho, Nieminen, Tommi & Pälli, Pekka (toim.). 2000. *Emergenssin kielelliset kasvot*. Folia Fennistica & Linguistica 24. Tampere: Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos.
Niiniluoto, Ilkka. 1983. *Tieteellinen päättely ja selittäminen*. Helsinki: Otava.
Niiniluoto, Ilkka. 1990. *Maailma, minä ja kulttuuri: emergentin materialismin näkökulma*. Helsinki: Otava.
Niiniluoto, Ilkka. 1999. *Critical Scientific Realism*. Oxford: Oxford University Press.
Nikanne, Urpo. 2015. What makes Conceptual Semantics special. Julkaisussa: Toivonen, Ida, Csuri, Piroska & Zee, Emile van der (toim.) *Structures in the Mind: Essays on Language, Music, and Cognition in Honor of Ray Jackendoff*. Cambridge (MA): MIT Press, 35–61.
Nikanne, Urpo. 2018. *Conceptual Semantics. A Micro-modular Approach*. Constructional Approaches to Language 23. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Nyman, Martti. 2000. Näkymätön käsi ja kielitieteen epistemologia. *Virittäjä* 104:1, 46–70.
Nyman, Martti & Suominen, Mickael. 2002. Ferdinand de Saussure: mies myytin takaa. Julkaisussa: Dufva, Hannele & Lähteenmäki, Mika (toim.) *Kielentutkimuksen klassikoita*. Soveltavan kielentutkimuksen teoriaa ja käytäntöä 4. Jyväskylä: Soveltavan kielentutkimuksen keskus, 9–66.
Phillips, Colin. 2009. Should we impeach armchair linguists. *Japanese/Korean Linguistics* 17, 49–64.
Pitkänen, Antti J. & Kohonen, Viljo. 1984. *Johdatus kvantitatiiviseen kielentutkimukseen ja alan atk-sovelluksiin*. Helsinki: Gaudeamus.
Plowright, David. 2011. *Using Mixed Methods: Frameworks for an Integrated Methodology*. Los Angeles (CA): Sage.
Popper, Karl. 1978. Three Worlds. The Tanner Lectures on Human Values, University of Michigan, April 1978. [verkkoaineisto]. [viitattu 5.2.2018]. Saatavissa: https://tannerlectures.utah.edu/_documents/a-to-z/p/popper80.pdf.
Quine, W. V. O. 1951. Two dogmas of empiricism. *The Philosophical Review* 60:1, 20–43.
Rosch, Eleanor. 1978. Principles of categorization. Julkaisussa: Rosch, Eleanor & Lloyd, Barbara (toim.) *Cognition and Categorization*. Hillsdale (NJ): Lawrence Erlbaum Associates, 27–48.
Rosch, Eleanor. 1999. Reclaiming concepts. *Journal of Consciousness Studies* 6:11–12, 61–77.
Saussure, Ferdinand de. 2002 [1916]. *Course in General Linguistics*. Illinois: Open Court.
Saussure, Ferdinand de. 2014 [1916]. *Yleisen kielitieteen kurssi*. [*Cours de linguistique générale*]. Toimittaneet Charles Bally ja Albert Sechehaye yhteistyössä Albert Reidlingerin kanssa. Suomentanut Tommi Nuopponen. Tampere: Vastapaino.
Scholz, Barbara C., Pelletier, Francis Jeffry & Pullum, Geoffrey K. 2016. Philosophy of linguistics. Julkaisussa: Edward N. Zalta (toim.) *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). [verkkoaineisto]. [viitattu 20.1.2020]. Saatavissa: https://plato.stanford.edu/archives/win2016/entries/linguistics/.
Schütze, Carson T. 2016. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Berlin: Language Science Press.
Searle, John R. 1980. Minds, brains, and programs. *Behavioral and Brain Sciences* 3:3, 417–457.
Searle, John R. 2008. Language and social ontology. *Theory and Society* 37:5, 443–459.
Seppänen, Lauri. 2004. *Kieltä ei ole. Uuspragmatistista kielifilosofiaa: Richard Rorty ja Donald Davidson*. Tampere: Tampere University Press.
Setälä, E. N. 1891. *Yhteissuomalainen äännehistoria*. 1. ja 2. vihko. Klusiilit. Spirantit. Nasaalit. Likvidat ja puolivokaalit. Helsinki: Suomalaisen Kirjallisuuden Seura.
Sherif, Muzafer. 1965. *The Psychology of Social Norms*. New York (NY): Octagon Books.
Sinha, Chris. 2014. Niche construction and semiosis: biocultural and social dynamics. Julkaisussa: Dor, Daniel, Knight, Chris & Lewis, Jerome (toim.) *The Social Origins of Language*. Oxford: Oxford University Press, 31–46.
Tieteen termipankki. 2018. [verkkoaineisto]. [viitattu 27.3.2018]. Saatavissa: http://www.tieteentermipankki.fi.
Tomasello, Michael. 2005. *Constructing a Language. A Usage-based Theory of Language Acquisition*. Cambridge (MA): Harvard University Press.

Tummers, José, Heylen, Kris & Geeraerts, Dirk. 2005. Usage-based approaches in Cognitive Linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory* 1:2, 225–261.

Töttö, Pertti. 2004. *Syvällistä ja pinnallista. Teoria, empiria ja kausaalisuus sosiaalitutkimuksessa*. Tampere: Vastapaino.

Varela, Francisco J., Kabat-Zinn, Jon, Rosch, Eleanor & Thompson, Evan. 2017. *The Embodied Mind. Cognitive Science and Human Experience*. Cambridge (MA): MIT Press.

Vogt, Paul W. 1999. *Dictionary of Statistics and Methodology. A Nontechnical Guide for the Social Sciences*. (2. p.) Thousand Oaks (CA): Sage.

Wang, Dahui, Li, Menghui & Di, Zengru. 2005. True reason for Zipf's law in language. *Physica A: Statistical Mechanics and its Applications* 358, 545–550.

Wettersten, John. 1992. *The Roots of Critical Rationalism*. Amsterdam: Rodopi.

Wiik, Kalevi. 1984. *Miksei* munoja *vaikka* kanoja*? Kantasuomen toisen tavun* a+i-*diftongin* a*:n kahtalaisen kehittymisen syiden pohdiskelua*. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 20. Turku: Turun yliopisto.

Wittgenstein, Ludwig. 1997 [1953]. *Philosophical Investigations*. Cornwall: Blackwell Publishing.

Wunderlich, Dieter. 1979 [1974]. *Foundations of Linguistics*. [*Grundlagen der Linguistik*]. Kääntänyt Roger Lass. Cambridge Studies in Linguistics 22. Cambridge: Cambridge University Press.

Zipf, George Kingsley. 1935. *The Psycho-biology of Language. An Introduction to Dynamic Philology*. Boston: Houghton Mifflin Co.

Zlatev, Jordan. 2012. Cognitive semiotics: An emerging field for the transdisciplinary study of meaning. *Public Journal of Semiotics* 4:1, 2–24.

# 4. Kielitieteen etiikka – hyvä tutkimus murroksessa

*Reetta Konstenius*

Tieteellisen tutkimuksen etiikka on nuori tutkimusala. Alle sadassa vuodessa se on kehittynyt yleispätevistä arvolauselmista alakohtaisiin normikokoelmiin, joista nyt ollaan siirtymässä digitaalisen ajan mikroetiikkaan. *Mikroetiikka* tarkoittaa tutkimuskohtaista eettisten huolenaiheiden tiedostamista ja tutkimukseen integroitua moraalifilosofista pohdintaa. Se on alhaalta ylös, datasta ja tutkimusasetelmasta kumpuavaa eettisesti motivoitua vuorovaikutusta niin tutkittavien kuin datankin kanssa (esim. ”Kokevatko osallistujat tämän virtuaalisen tilan yksityisenä vai julkisena?”). *Makroetiikaksi* nimitetään ylhäältä alas kategorisesti sovellettavia yleisiä tutkimuseettisiä toimintakoodeja. Tällöin sovelletaan klassisia tutkimuseettisiä normistoja, joita yliopistot, tutkimuseettiset neuvottelukunnat tai alan yhdistykset ovat kirjanneet (esim. ”Älä vääristele, valehtele tai sepitä tutkimustuloksia.”).

Tässä luvussa tarkastellaan ensin *etiikan* käsitettä ja laajan vastuun etiikan historiallista kehitystä. Tämän jälkeen käsitellään varsinaista tutkimusetiikkaa, käsityksiä ihmisoikeuksista ja kielitieteen suhdetta makroeettisiin koodeihin. Lopuksi selvitetään verkkoon kohdistuvan kielitieteen ja teknologian aiheuttamia muutostrendejä tutkimusetiikassa.

Tämä luku ei tarjoa kaikkialla pätevää tutkimuseettistä normistoa kielentutkijan arkipäivään, sillä tutkimusta muiden kuin biolääketieteellisten alojen eksperimenttien etiikasta on olemassa varsin

rajallisesti (ks. yleisesti Clarkeburn & Mustajoki 2007). Tämä johtuu siitä, että etiikantutkimus on yleisesti nuori ala, ja kielitieteen etiikantutkimus on vielä sitäkin nuorempaa. Kysymys siitä, koskeeko humanistisiakin aloja tutkimuseettinen koodisto, esitettiin oikeastaan vasta 2000-luvulla (ks. Hallamaa, Launis, Lötjönen & Sorvali 2006). Kun nykyisin Suomessa julkaistaan vuosittain keskimäärin noin kymmenen biotieteiden etiikkaa koskevaa väitöskirjaa (ks. Länsimies-Antikainen 2008) ja alaa voi opiskella monessa yliopistossa ulkomailla, kielitieteen etiikan väitöskirjoja ei toistaiseksi ole juurikaan julkaistu enempää Suomessa kuin ulkomaillakaan. Erot eri aloilla tehdyssä etiikan tutkimuksessa johtavat epäsuhtaan etiikkaa koskevan tiedon laadussa ja määrässä (Konstenius 2014).

Tässä luvussa esiin tulevat huomiot kielitieteen etiikasta ovat peräisin erilaisista kenttämetodioppaista sekä kielitieteellisten yhdistysten ja julkaisujen omilta internetsivuilta. Suomessa ainoaa kielitieteen etiikkaa käsittelevää väitöskirjaa, Konsteniuksen (2014) teosta *Empiria, eksperimentti ja etiikka*, on tässä luvussa referoitu runsaasti. Kirjallisuuden niukkuus implikoi sitä, että alalla ei ole syvällistä julkista ja vakiintunutta tutkimuseettistä pohdintaa. Siksi keskustelua kielitieteen etiikan koodistosta ei voi tässä referoida. Tavoitteena onkin nyt tuoda esille erilaisia puolia, näkemyksiä ja ongelmia kielitieteen etiikassa siten, että kielentutkimuksen yhteisö voi kollektiivisesti – jokainen oman tutkimuksensa osalta – vastata tutkimuseettisiin kysymyksiin. Juuri tämä onkin nykyisen mikroeettisen tutkimuseettisen suuntauksen sanoma: valmiiden vastausten sijaan tutkijan on aktiivisesti itse pohdittava ja harjoitettava omaan tutkimustyöhönsä kohdistuvaa moraalifilosofiaa.

## 4.1. Kielentutkimuksen etiikka informaatioyhteiskunnassa

Internet on aiheuttanut yhteiskunnallisen murroksen, jota filosofit ovat sen vallankumouksellisuudessa verranneet painokoneen keksimiseen viisisataa vuotta sitten. Uudet teknologiat ovat mahdollistaneet uusia viestintäkeinoja, ja tällä on pitkälle meneviä kerrannaisvaikutuksia, joita emme vielä kaikkia tunne. Osa muutoksista vaikuttaa myös kielitieteen etiikkaan.

Kielitieteen etiikkaa on joiltain osin mielekästä tarkastella juuri sen teknologian kautta, johon kieli on sidottu ja jolla kieltä välitetään tai tutkitaan. Puhuttua, painettua tai sähköisesti internetistä koottua kieltä koskevat tutkimuksessa erilaiset tutkimuseettiset normit. Tämä johtuu siitä, että kielen sidos ihmisten oikeuksiin on näissä eri tilanteissa tulkittu erilaiseksi.

Tutkimusmetodi voidaan tässä kontekstissa ymmärtää tekniikkana, työkaluna, jolla uutta tietoa muodostetaan reaalitodellisuudesta. Tässä luvussa käsitellään niitä yleisiä normeja, jotka ohjaavat metodin eettistä käyttöä. Kuten oikeiden työkalujen myös metodien välillä on suuria eroja, joten niiden aiheuttamat potentiaaliset vahingotkin eroavat toisistaan. Siksi niiden käyttöä ohjaavat normistot ovat myös erilaisia. Samalla tavalla kuin moottorisahan ja ruuvimeisselin käytössä turvajärjestelyjen määrä ja laatu poikkeavat toisistaan, niin eroavat myös tutkimustekniikoiden turvallista käyttöä ohjaavat normistot. Näistä metodien erityispiirteistä keskustellaan eri metodien yhteydessä tämän teoksen eri luvuissa.

Informaatioyhteiskunnalle on ominaista, että suuri osa sen keskeisiä toimintoja – talous sekä sosiaalinen ja poliittinen elämänpiiri – on sidoksissa teknologiaan. Nykytutkimus voidaan siksi jakaa online- ja offline-tutkimukseen. Online-tutkimusta on kaikki se tutkimus, joka suoraan tai epäsuorasti on sidoksissa teknologian medioimaan kieleen, teknologian luomaan virtuaaliseen tilaan tai toimintaan (*online* 'tietokone, muu teknologinen väline tai ihminen, joka on yhteydessä tietoverkkoon, yleensä internetiin'). Offline-tutkimusta on yleensä traditionaalinen internetaikaa edeltävä kielentutkimus (*offline* 'tietokoneen, muun teknologisen välineen tai ihmisen kytkeytymättömyys tietoverkkoon, yleensä internetiin').

Humanistisilla sekä yhteiskunta- ja käyttäytymistieteellisillä aloilla offline-tutkimusetiikka jaetaan yleensä kahtia: ihmiseen kohdistuvan tutkimuksen ja tekstin tutkimuksen etiikkaan. Kun tutkimus kohdistuu tekstiin, normisto eroaa huomattavasti niistä säännöistä, joita noudatetaan ihmistutkimuksessa (engl. *human subject research*). Tämä kahtiajako ei näytä kuitenkaan toimivan suoraviivaisesti sähköisissä ympäristöissä. Teknologian välittämässä ympäristössä tutkimus kohdistuu yleensä

kirjoitettuun kieleen. Kohdistuuko esimerkiksi blogin kommenttien tutkimus tekstiin vai ihmiseen? Eri tutkijat näyttävät olevan tästä eri mieltä, minkä vuoksi he soveltavat erilaisia eettisiä koodeja. Kuitenkaan samaan ihmiseen, samaan tekstiin tai samaan ilmiöön ei voi soveltaa kahta eri normistoa sen mukaan, mikä on tutkijan oma taustateoria ja käsitys. Ihmisen oikeudet eivät voi olla ”katsojan silmässä”, vaan niiden tulee olla ihmisen pysyvä ominaisuus. Tällä hetkellä digitaalisten ilmiöiden olemuksesta käydään vilkasta filosofista keskustelua. Esimerkiksi historian alan yhdistykset ovat esittäneet, että kaikkia sähköisissä ympäristöissä esiintyviä ilmiöitä käsiteltäisiin teksteinä. Tämä mahdollistaisi sen, että tutkijan ei tarvitsisi erikseen selvittää, onko ihminen vai kone automaattisesti tuottanut tekstiä (esim. urheiluotteluiden tulosraportit). Täysin vastakkaisiakin käsityksiä on esitetty: kaikki sähköisesti välitetty kieli kohdistuu viime kädessä ihmiseen.

On huomattu, että on olemassa suuri tarve määrittää yhdet yhteiset pelisäännöt, joilla toimia sähköisissä ympäristöissä. Sääntöjen tulisi olla sellaiset, että myös liikeyrityksissä tehtävä tutkimus (nk. sale, ks. Johdannon luku 2.2. tk.) olisi akateemisen tutkimuksen kanssa samalla linjalla. Ei ole erityisiä moraalisia syitä, joiden vuoksi osakkeenomistajien tai harvojen yksityistä etua ajavien tutkimusta ei voisi rajoittaa sama tiukka säännöstö, joka rajoittaa myös akateemista tutkimusta.[1]

Teknologian nopea kehitys, teknologian välittämä diversiteetti, jossa kieli esiintyy, ja myös tutkijoiden käyttämät useat erilaiset sähköiset työkalut ovat saaneet monet päättelemään, että yleiset makroeettiset, ylhäältä alas kategorisesti sovellettavat koodit eivät enää sovellu tutkimuksiin (ks. esim. de Costa 2015). Uudet teknologiat ovat saaneet aikaan mikroeettisten[2] lähestymistapojen yleistymisen: tutkimuksen kuluessa vastataan eettisiin kysymyksiin, ja tutkimusasetelma sekä tutkimuseettiset päätelmät ovat monesti ainutlaatuisia, tapauskohtaisia ja yksilöllisiä.

Eräät ihmistieteiden filosofit, kuten Hans-Georg Gadamer (1976), Michel Foucault (1982) ja nykyaikaisen feministisen tiedekritiikin edustajat, ovat sitä mieltä, että arvovapaata tiedettä ei ole olemassa. Etenkin Julia Kristeva (1988) ja Luce Irigaray (2017) ovat kritisoineet objektiivisena ja neutraalina esiintyvää tieteen kieltä, joka kuitenkin kaiken muun inhimillisen toiminnan tavoin perustuu valintoihin, intresseihin

ja pyrkimyksiin. Mikroeettinen trendi, joka pakottaa tutkijan pohtimaan omia arvovalintojaan tutkimuksessa, on tuonut siten muutoksen tähän jännitteeseen. Moraalifilosofinen pohdinta osana tutkimusta ja arvoihin perustuva argumentaatio ovat tulleet jäädäkseen. Hyvään tutkimukseen kuuluu nykyään monipuolinen arvojen, eettisten toimintakoodien, johtamistapojen ja lähtöoletusten kyseenalaistaminen. Eettinen pohdinta kyseenalaistaa tutkijan omankin aseman, intressit ja ennakkokäsitykset. Tutkimusjulkaisuihin integroitu moraalifilosofinen pohdinta tuo siten suuremmassa määrin esille tutkijan subjektiivisuutta ja erehtyvyyttä. Eettinen positio on välttämättä avoin muiden inhimillisille kokemuksille, tulkinnoille ja näkemyksille. Kielentutkimuksen soveltaminen digitaalisessa ympäristössä tuo siten mukanaan muutoksia, ei vain tieteen etiikkaan vaan myös tieteelliseen argumentaatioon.

## 4.2. Etiikan lajit ja tutkimus

Käsite *etiikka* viittaa niihin toimintakoodeihin, jotka on johdettu arvoista. Etiikan eli moraalifilosofian tutkimusaluetta ovat siten arvot ja se tapa, jolla arvoista johdetaan oikeaa toimintaa ohjaavia sääntöjä tai normeja. Hyvää (tai oikeaa) on sellainen, joka lisää tai säilyttää itseisarvoa, ja pahaa (tai väärää) vuorostaan sellainen, joka tuhoaa tai vähentää itseisarvoa.

**Itseisarvot** ovat päämääriä. Yleensä esimerkiksi ihmiselämää pidetään sellaisena itseisarvona, jota ei voi perustella muilla tekijöillä ja jota kannattaa tavoitella sen itsensä vuoksi. **Välinearvoilla** on arvoa, koska niiden avulla saavutetaan itseisarvoja. Näin esimerkiksi tutkimusetiikalla voisi ajatella olevan välinearvoa, koska sillä tutkimuksen kuluessa turvataan ihmisten hyvinvointi.

Toimintakoodien johtamisen tavat voi karkeasti jakaa kolmeen: seurausetiikkaan, velvollisuusetiikkaan ja hyve-etiikkaan. **Seurausetiikka** määrittää teon moraalisen arvon sen mukaan, missä määrin teon seuraus palvelee itseisarvoja. Seurausetiikan tunnetuin muoto on utilitarismi eli **hyötyetiikka**. Sen mukaan eettisen toiminnan johtava periaate on, että aina on tehtävä se teko, josta on kaikkien ihmisten kannalta suurin

mahdollinen hyöty. Tämä tarkoittaa, että jokin normaalisti eettisesti tuomittava teko voi jossain tilanteessa olla moraalinen. Jos esimerkiksi kielentutkija rikkomalla yhtä pienempää sääntöä, esimerkiksi vaitiolovelvollisuutta, turvaisi koko kieliyhteisön etniseltä puhdistukselta, olisi se utilitaristisen periaatteen mukaan hyväksyttävä pieni paha, jotta saavutettaisiin laajempi hyvä.

**Velvollisuusetiikassa** johdetaan arvosta kategorinen sääntö. Siksi jotkin teot ovat aina kiellettyjä tai velvoittavia riippumatta kontekstista. Velvollisuusetiikassa voidaan tehdä sääntöluettelo tai yksi yleinen sääntö. Sääntöluettelossa jokaisesta erillisestä arvosta johdetaan koodi. Jos esimerkiksi totuus käsitetään itseisarvoksi, on vastattava totuudenmukaisesti silloinkin, kun tietää vastauksen johtavan omaan tai toisen tuhoon. Juuri totuudenmukaisuuden katsotaan kuuluvan tutkijan velvollisuuksiin – tutkijan ei katsota koskaan voivan antautua vilppiin, valheeseen tai vääristelyyn.

Yleensä tutkimuseettisissä koodistoissa ja kansallisvaltioiden lakinormistoissa on määritelty arvohierarkioita: esimerkiksi ihmiselämä menee kategorisen totuuksien julistamisen edelle. Jos tutkija esimerkiksi tekee töitä etnisen puhdistuksen kohteena olevan vähemmistön kanssa, hänen tulee harkita erityisen huolella, mitä tosiasioita hän voi tästä ryhmästä julkaista. Tutkittavien henkilötiedoista vaikenemista edellytetään suurimmassa osassa tutkimustyyppejä.

Velvollisuusetiikka voi perustua myös yhteen ainoaan periaatteeseen, josta muut toimintanormit johdetaan. Tunnetuin tällaisista velvollisuuseettisistä normeista on Kantin kategorinen imperatiivi. Se kehottaa miettimään, onko teko, jonka on päättänyt tehdä, moraalisesti niin hyvä, että siitä voi tehdä maksiimin eli toivoa sen olevan yleispätevä, kaikkia yhtäläisesti koskeva laki. Useat kenttätutkimusoppaat kehottavatkin pohtimaan uusia tilanteita, joihin ei ole olemassa ennalta eettistä koodia kantilaisittain: ”toivoisinko, että esimerkiksi tästä käyttämästäni tavasta tulisi sellainen yleinen laki, joka koskisi myös itseäni ja lapsiani”. Kategorista imperatiivia kutsutaan myös *kultaiseksi säännöksi*, sillä sen katsotaan ohjaavan ihmisen moraalifilosofista päättelyä oikeaan suuntaan hankalissakin kulttuurisidonnaisissa tutkimustilanteissa. Esimerkiksi Karen Rice (2004) ehdottaa, että tutkija voi käsitteellistää

minkä tahansa aspektin tutkimusmetodistaan kysymyksenä, miten siinä suhtaudutaan tutkimukseen osallistuviin ihmisiin, ja samalla pohtia, tahtoisiko itseään kohdeltavan näin.

Ihminen hahmottuu osaksi tutkimusta ainakin seuraavilla tavoilla:

- tutkimuksen tekeminen ihmisillä (*on humans*)
- tutkimuksen tekeminen ihmisten kanssa (*with humans*)
- ihmiset tutkimuksen välineinä (*humans as means to data or results*)
- tutkimuksen tekeminen ihmisille (*for humans*).

Ihmisen käyttö välineenä ei ole yleisesti hyväksyttävää, joten kielentutkimusta ohjaavana periaatteena tulisi olla, että tutkimusta tehdään ihmisille ja ihmisten kanssa. Tutkimuksen eettinen asennoituminen yhteisöön ja sen jäseniin näkyy Ricen (2004) mukaan tutkimuksen kuluessa tehdyissä metodologisissa ratkaisuissa.

**Hyve-etiikan** näkökulma on hyvin toinen kuin seuraus- ja velvollisuusetiikan. Se keskittyy toimijan omiin ominaisuuksiin, tunteisiin ja motivaatioihin. Hyve-etiikka korostaa ihmisen sisäisen motivaation merkitystä ja toimijan omaa käsitystä hyvästä ja pahasta. Hyve-etiikka hahmottaa eettiset valinnat ihmisyyden osana, kunkin ihmisen luonteenpiirteenä, joka toisaalta selittää tämän ihmisen käytöstä (antelias ihminen antaa omastaan) ja toisaalta on ihmisen omassa kehityksessä tavoittelemisen arvoinen piirre (ihminen haluaa olla vähemmän itsekäs). Hyve-etiikka pyrkii määrittämään ne ominaisuudet, joita hyveellisellä ihmisellä on. Ammattietiikka on usein eräänlaista hyve-etiikkaa. Hyvää tutkijaa luonnehditaan usein sanoilla *tarkka*, *rehellinen*, *lahjomaton* ja *utelias*. Nykypäivän uusliberaalissa hengessä tutkijalta vaaditaan myös kommunikaatiotaitoja, ryhmähenkeä ja tuloksellisuutta (Niiniluoto & Sihvola 2005).

Tutkimuseettiset normistot ovat yleensä yhdistelmiä edellä kuvatuista kolmesta tavasta johtaa eettisiä toimintakoodeja. Kansallisvaltioiden lainsäätäjät ja oikeuslaitokset, tutkimusalojen yhdistykset (esim. BAAL 2006) ja tiedelehdet (esim. *TESOL Quarterly*) kehittävät tutkimuseettisiä normistoja. Toisinaan myös tiettyihin erityisryhmiin (esim. lapset), kieliryhmiin (vähemmistökieliryhmät) tai kieliin (uhanalaiset kielet) liittyy

omia normistoja. Esimerkiksi yhdistys Applied Linguistics Association of Australia on julkaissut vuonna 1998 ohjeiston (ALAA 1998), jossa määritellään, miten aboriginaalien oikeudet turvataan tutkimuksen kuluessa. Samankaltaisen ohjeiston, jossa suojataan vähemmistöjä, on julkaissut myös yhdistys American Association for Applied Linguistics (AAAL 2011) vuonna 2011. Ero näiden kahden ohjeiston välillä on, että toinen perustuu etnisyyteen ja toinen kielelliseen kriteeriin: AAAL suojaa kaikki korostuksella (engl. *accented speech*) englantia puhuvat vähemmistöryhmät.

Suomessa Tutkimuseettinen neuvottelukunta[3] ohjaa tutkimuseettistä keskustelua ja on julkaissut oppaita (esim. Hyvä tieteellinen käytäntö 2012) sekä verkkosivun vastuullinentiede.fi. Vastuullinen tiede -sivusto opastaa yleistajuisesti tutkimusetiikan hyviin käytänteisiin, ja se sisältää tietoa myös tiedeviestinnän ja avoimen tieteen periaatteista.

Tutkimuseettisten normistojen välillä on toisinaan ristiriitoja, joten tutkijan kannattaa tutustua perinpohjaisesti oman alansa eksplisiittisesti kirjoitettuihin normistoihin ja erilaisiin käsityksiin hyvästä tutkimuksesta. Tutkimusetiikassa tapahtuu myös nopeita muutoksia uusien metodien ja teknologioiden kehittyessä. Esimerkiksi vuonna 2018 voimaan tullut uusi EU:n tietosuoja-asetus asettaa joitakin uusia rajoituksia tavoille kerätä tietoa sähköisissä ympäristöissä. Jos omalla alalla huomaa ristiriitoja eri normistojen välillä tai jotkin normit vaikuttavat liian leväperäisiltä tai aivan liian ankarilta, nämä erilaiset huomiot tulee tuoda esille tutkimusjulkaisussa ja keskusteluttaa oman alan tiedeyhteisössä. Eettinen normisto on eettinen vain sikäli, kuin se tosiasiassa suojaa tutkimukseen osallistuneiden henkilöiden oikeuksia. Tutkimusetiikka pelkkänä **hyveen viestittämisen** (engl. *virtue signaling*) välineenä on eräs korruption muoto (Orlitzky 2017).

## 4.3. Laajan vastuun etiikka

Nimensä mukaisesti laajan vastuun etiikka pohtii tutkimuksen laajempia yhteiskunnallisia, historiallisia ja ekologisia vaikutuksia. Mitä seurauksia tutkimuksen tuloksilla voi olla? Voiko tutkimuksen tuloksilla

saada tahattomasti aikaan vahinkoa, harmia tai haittaa yksilöille tai yhteisöille? Onko tutkimuksella globaaleja vaikutuksia? Ottaako se tulevien sukupolvien oikeudet huomioon? Tulisiko tutkimusta tehdä lainkaan? Koska tutkijanetiikka kieltää vääristelemästä tutkimustuloksia, ainoa vaihtoehto on jättää tekemättä tai julkaisematta sellainen tutkimus, joka saattaisi aiheuttaa vahinkoa.

Entä kuinka vakava potentiaalisen vahingon tulee olla, jotta tutkijalla on velvollisuus jättää tutkimus julkaisematta? Mitä eroa on atomipommilla ja kieliryhmän syrjimisellä? Tässä yhteydessä on mielekästä tuoda esiin, miten pohdinta tutkijan laajasta vastuusta alkoi. Niin kutsuttu Manhattan-projekti sysäsi liikkeelle tapahtumaketjun, joka sai alkunsa, kun natseja paennut unkarilainen fyysikko Leo Szilard sai Albert Einsteinin vakuuttuneeksi siitä, että Otto Hahnin saamia tuloksia atomin halkaisemisesta voisi soveltaa sodankäynnissä. He kirjoittivat yhdessä presidentti Rooseveltille kirjeen, jossa vakuuttivat tämän atomipommia koskevan tutkimuksen tarpeellisuudesta. Oli kehitettävä atomipommi ennen natsi-Saksaa.

Kysymys tutkijoiden laajasta vastuusta nousi, kun suuri yleisö halusi selvittää, mikä oli tutkijoiden moraalinen vastuu, kun Yhdysvallat lopulta pudotti atomipommit vuonna 1945 sodassa Japania vastaan sellaisessa tilanteessa, jossa Saksa ei enää ollut sotilaallinen uhka. Oli myös jo käynyt ilmi, ettei Saksalla edes ollut atomipommia. Manhattan-tutkimusprojekti johti satojen tuhansien siviilihenkilöiden kuolemaan. Olivatko tutkijat ja atomipommin kehittäjät siis moraalisessa vastuussa tapahtuneesta? Kysymys tutkijoiden yleisestä moraalisesta vastuusta asetettiin tämän jälkeen uudelleen.

Julkisuudessa Manhattan-projektin tutkijat esittivät, etteivät he tienneet, että pommilla voisi olla tuhoisia vaikutuksia. He siis vähättelivät moraalista vastuutaan vetoamalla tiedolliseen argumenttiin. Moraalifilosofiassa on vaikeampaa perustella vastuuta, jos tekijällä ei ole ollenkaan tietoa tekonsa todellisesta luonteesta ja sen seurauksista. On kuitenkin epäuskottavaa, että tutkijat eivät ymmärtäisi, että pommilla voi pommittaa. Kävikin ilmi, että atomipommin vaikutuksia testattiin Yhdysvalloissa etukäteen, myös ihmisillä (Roelcke & Maio 2004). Esimerkiksi Vanderbiltin yliopistossa 800:lle raskaana olevalle naiselle

annettiin radioaktiivista rautaa, jotta saatiin selville radioaktiivisuuden vaikutus sikiöön (Eaton & Kennedy 2007, 39). Tutkijat siis tiesivät atomipommin seurauksista ihmisille ja osallistuessaan tutkimukseen myös hyväksyivät sen, että tutkimuksen kuluessa vahingoitettiin oman maan kansalaisia, naisia ja lapsia.

Vaikka nykypäivänä tällaiset toisen maailmansodan aikaiset tapahtumat tuntuvat ehkä kaukaisilta ja niiden pohtiminen voi vaikuttaa tarpeettomalta, liittyy tapahtumien taustalla oleviin kysymyksenasetteluihin ja oletuksiin ajankohtainen piirre, jota on syytä pohtia myös kielitieteessä. Kuten tuolloin niin myös nyt juuri uudet teknologiat luovat sellaisia uusia mahdollisuuksia ja toimintoja, joiden yhteydessä vanhat oletukset moraalista ja moraalisesta toiminnasta on tarkistettava. Juuri vanhat yleisesti hyväksytyt moraaliset näkökannat nimittäin mahdollistivat laajamittaisen tuhon jättäessään katveeseen suoraan totuttuihin periaatteisiin sopimattomat uudet ilmiöt. Manhattan-projektin mahdollinen vastine kielitieteessä voisi olla vaikkapa se, millä lailla kielentutkijat, jotka osallistuisivat esimerkiksi kielellistä profilointia suorittavan ohjelmiston tekoon, olisivat moraalisesti vastuussa sen käytöstä ja seurauksista, jos ohjelmistoa käytettäisiin vaikkapa vakaumukseltaan tietyn ihmisryhmän vainoamiseen ja tuhoamiseen.

Toisena esimerkkinä voi esittää lääkärin ammattietiikan kehittymisen. Hippokrateen vala ja sitä seurannut tutkimuseettinen kirjallisuus (esim. Percival 1803) pitivät lääkärin oikeutta toteuttaa lääketieteellisiä toimenpiteitä potilaalle ensisijaisena suhteessa potilaan oikeuksiin. Jos lääketieteellisen toimenpiteen tekeminen vaatisi valehtelemista tai totuuden peittelyä potilaalta, lääkäri olisi suorastaan velvollinen valehtelemaan. Tämä käsitys oli yleisesti niin hyväksytty, että vaikka 1920- ja 1930-luvuilla sekä Yhdysvalloissa että Euroopassa jotkut harvat ehdottivat tutkimuseettistä normistoa ja potilaiden hyväksynnän pyytämistä tutkimuksiin, valtaosa tiedeyhteisöstä piti tällaista rajoittavaa koodistoa tarpeettomana ja jopa tieteen hengen vastaisena. Tieteellinen yhteisö piti siis oikeutenaan tehdä kokeita haavoittuvaisilla ryhmillä, kuten lapsilla, psykiatrisilla potilailla, värillisillä ja köyhillä. Manhattan-tapaus ja natsi-Saksan keskitysleirien epäinhimilliset kokeet olivat tämän laajasti hyväksytyn asenteen ääri-ilmentymiä. Vasta järkyttävien tapahtumien

kautta suuri yleisö tuli tietoiseksi siitä, että tarvitaan yleiseettiset normistot, jotka takaavat, että raskaana oleva nainen voi mennä synnytyssairaalaan eikä hänen tarvitse pelätä, että häneen ruiskutetaan hänen tietämättään myrkkyä. Hän voi myös lähettää lapsensa logopedille ja olla täysin varma, että siellä ei testata salaa sähköshokkien vaikutusta ääntämiseen.

Vallitseva käsitys viime vuosisadalla siis oli, että tieteen tekijöillä on erityisiä oikeuksia pyrkiessään tietoon, mikä näkyi edellä esitetyn lisäksi muun muassa alkuperäiskansojen ja kielten tutkimuksessa. Antropologisessa kielitieteessä ja muussa kenttätutkimuksessa rokottamattomien alkuperäiskansojen pariin mentiin välittämättä seurauksista. Länsimaisten lähetyssaarnaajien ja tutkijoiden tuomat virukset ovatkin tuhonneet kieliyhteisöjä mittakaavassa, jota esimerkiksi Tierney (2000) on verrannut kansanmurhaan.

Einstein, joka katui osallisuuttaan atomipommin kehityksessä, oli yksi ensimmäisistä, jotka ehdottivat tutkijan etiikan täyden vastuun mallia. Tämän mallin perusajatus on, että voi olla olemassa sellaisia asioita, joita ihmiskunnan ei tulisi tutkia, ja esineitä, joiden valmistusohjetta ihmisillä ei tulisi olla. Tieteellisellä tiedolla voi olla yhteisön näkökulmasta epätoivottavia seurauksia, tai tieteellisen tutkimuksen tulos voi luonteeltaan olla sellainen, että se loukkaa keskeisiä arvoja. Tällaisia ovat esimerkiksi ihmisen kloonaus sekä biologiset ja autonomiset aseet. Kielentutkimuksessa vastaavanlaisia alueita voivat olla ihmisten yksityisyyteen ja autonomisuuteen kajoavat digitaaliset profilointi- ja manipulointiohjelmat. Täyden vastuun malli tarkoittaa, että kullakin tutkijalla on vastuu toimia moraalisesti, myös silloin kun vallitseva eettinen koodisto ei vielä ole virallisesti rajoittanut tieteellistä tutkimusta. Tieteessä tutkijan voidaan aina ajatella olevan uuden tiedon äärellä; vanhan tiedon toistaminen yksinomaan ja sellaisenaan ei ole tieteen tavoite. Loogisesti vanhat eettiset koodistot eivät voi sisällyttää itseensä sitä, mitä vielä ei tiedetä. Yksittäisen tutkijan vastuu on siten aina pohtia tutkimuksensa seurauksia: Einsteinin mukaan tutkija on vastuussa siitä tiedosta, jonka hän saattaa yhteisön ulottuville.

Myös toisenlaisia näkemyksiä on esitetty. Esimerkiksi Wiberg (2006, 265) esittää, että tutkijaa ei voi laittaa yksin vastuuseen muiden teoista,

vaikka ne perustuisivatkin hänen tutkimuksensa tuloksille. Tämän näkemyksen mukaan tutkija ei kanna muiden toimijoiden moraalista taakkaa yksinkertaisesti siksi, ettei hänellä ole valtaa määrätä heidän tekojaan. Tiedon jakaminen ei implikoi vastuuta sen käytöstä.

Graham (1978), joka on paljon pohtinut laajaa etiikkaa, on kuitenkin eri mieltä ja esittää, että tutkija on aina moraalisesti vastuussa tutkimuksensa seurauksista. Hän antaa esimerkin: Natsi-Saksan vallan aikana tutkija löytää perinnöllisen Tay-Sachsin taudin juutalaisten perimästä. Koska natsi-Saksan poliittisessa kontekstissa tiedon julkaisemisen seuraukset voidaan arvioida ja tutkija siten voi tehdä moraalisen valinnan, on hän myös vastuussa siitä, että hänen tutkimustaan käytettäisiin vähemmistön etnisessä puhdistuksessa (ks. myös Lötjönen 1999; Konstenius 2014).

Wiberg ja Graham ehkä puhuvat eri mittakaavoissa, ja siksi päätelmät eroavat. Laajan vastuun etiikan keskustelussa on mielekästä pohtia, olisiko oma tutkimus vahingon väline, välittäjä tai sen aiheuttaja. Tuhon ja vahingon välineitä voivat olla autot, tietokoneet ja muut erilaiset arkiesineet, joita ilman emme ehkä voisi nyky-yhteiskunnassa toimia. Autonomiset aseet tai vainottujen ryhmien automaattiseen kielen tunnistamiseen kehitetyt välineet taas ovat välittömiä vahingon aiheuttajia: niiden sovellusalue on vahingon aiheuttaminen.

Kielitieteessä ja antropologiassa Grahamin esimerkkiä on usein käytetty havainnollistamaan yhteiskunnallisen tiedostamisen merkitystä. Julkaistua tietoa voi mahdollisesti käyttää haavoittuvaista ryhmää vastaan. Esimerkiksi Dorian (1993) painottaa, että tutkimus tehdään aina tietyssä historiallisessa ja poliittisessa kontekstissa,[4] jota kielentutkijan tulee pohtia kieliyhteisön etujen näkökulmasta. Myös Linguistic Society of America -yhdistyksen[5] virallisten ohjeiden mukaan kielentutkijan tulee arvioida tutkimuksensa mahdollisia sosiaalisia ja poliittisia seurauksia.

Pietarisen ja Launiksen (2002) mukaan tutkijan on kehitettävä ennakoivaa näkemystietoa siitä, miten hänen tutkimuksensa tulokset voivat vaikuttaa. Vaikka tulevaisuus on arvaamaton eikä tutkija voi sitä kontrolloida, tutkija ei vapaudu eettisestä vastuusta silloin, kun hän itse uskoo, että tutkimustuloksia voidaan käyttää keskeisiä ihmisarvoja loukkaavalla tavalla. Pietarisen näkemyksen mukaan tutkijan velvollisuuksiin kuuluu kehittää näkemystietoa oman tutkimuksensa mahdollisista

vaikutuksista. Näkemystiedon kehittäminen kuuluu siis hänen mielestään tutkijan ammattitaitoon ja ammattietiikkaan.

Näkemystiedon kartuttamiseksi tutkimus tulisi aloittaa eettisillä kysymyksillä. Creswellin ja Creswellin (2017) mukaan tutkijan tulee pohtia esimerkiksi sitä, mikä tutkimuksen tarkoitus on. Jos tutkija ei keksi mitään funktiota tutkimukselleen ja näkee sen turhana, silloin se on yhteiskunnallisten varojen haaskausta ja hänen tulisi harkita jotakin muuta tutkimuskysymystä. Jos taas tutkimuksella on ainakin yksi merkittävä anti, silloin sillä on myös seurauksia. Tätä seurausta ja sen yhteiskunnallisia sidoksia tutkijan on pohdittava eri näkökulmista. Tällainen tutkimuseettinen pohdiskelu koskee niin perus- kuin soveltavaakin tutkimusta.

Wolfram (1993, 227) on muotoillut ”kielitieteellisen kiitollisuusperiaatteen”. Sen mukaan on mukaan kielentutkija on aina kiitollisuudenvelassa sille yhteisölle, jolta on kerännyt datan, ja on näin ollen velvollinen myös toimimaan yhteisön parhaaksi (ks. Lingvistinen etnografia toimintayhteisöissä; Kielten dokumentointi ja kieliopin kuvaus; Antropologinen lingvistiikka tk.). Siksi kielentutkimuksessa tulisi hänen mukaansa aluksi pohtia tutkimusongelmaa ja tutkimusasetelmaa tutkittavan kieliryhmän etujen perspektiivistä. Onko tutkimukseen valittu käsitteistö tutkittavan yhteisön identiteetin ja etujen mukainen? Millaisessa sosiopoliittisessa kontekstissa tutkimusta tehdään? Voiko tutkimuksen julkaisemisesta koitua haittaa ryhmälle? Miten siitä koituisi etua? Wolframin (mt.) näkemys tutkijan eettisistä velvollisuuksista on siten Pietarisen ja Launiksen (2002) käsityksiä vaativampi: tutkijan ei ainoastaan tule varoa tekemästä haittaa, vaan hänen tulee tutkimuksellaan pyrkiä saamaan aikaan hyviä asioita kieliyhteisölle.

Joskus tutkimuksen päämäärä itse asiassa on vahingoittaa tai heikentää tutkittavaa ilmiötä. Rikollisuutta, korruptiota, ihmiskauppaa tai köyhyyttä tutkivan on yleensä moraalista asettaa kysymyksensä siitä näkökulmasta, miten ilmiötä vähennetään, eikä suinkaan siitä, miten sitä lisättäisiin. Kielitieteessä tällaisia päämäärähakuisia tutkimusalueita ovat esimerkiksi seksistinen kieli, propaganda tai vähemmistön ihmisarvoa alentava terminologia. Vahingoittamisen kieltävän imperatiivin seuraaminen sokeasti olisi epäeettistä tässä kontekstissa. Laajan vastuun

eettisten kysymysten luonteeseen kuuluukin, että laajan vastuun pohdintaa ei voi mekaanisesti soveltaa ja kopioida muista tutkimuksista. Uusi tutkimus on usein ainutlaatuinen, ja sillä on muuttuvassa maailmassa uusia suhteita, joita ei aikaisemmissa tutkimuksissa ole ollut (Graham 1981; Löppönen 1991; Wiberg 2006; Pietarinen & Launis 2002). Laajan vastuun etiikka edellyttää yhteiskunnallista – ja nykyään myös globaalia ja teknologista – valveutuneisuutta. Uudet teknologiat mahdollistavat uusia tapoja vahingoittaa tai edistää erilaisten kieliyhteisöjen hyvinvointia.

## 4.4. Tutkimusetiikka

Toisen maailmansodan jälkeen tieteellisessä yhteisössä syntyi yksimielisyys siitä, että tieteellistä tietoa ei ole oikeutta hankkia millä hinnalla hyvänsä. Ihmisarvojen ja eettisten normien on ohjattava tutkimusasetelmien suunnittelua ja rajoitettava vahinkoa aiheuttavien tekniikoiden käyttöä. Natsi-Saksassa tehdyt ihmisyyttä vastaan rikkovat lääketieteelliset kokeet olivat nousseet Nürnbergin oikeudenkäynnin kautta yleiseen tietoisuuteen.[6] Myös YK:n yleismaailmalliset ihmisoikeudet kirjoitettiin, jotta historia ei toistaisi itseään. YK:n yleinen ja universaali ihmisoikeusjulistus määrää kaikille ihmisille yhtäläiset oikeudet; artiklassa 7 säädetään erityisesti ihmisen oikeudesta itse päättää siitä, osallistuuko hän tieteelliseen tutkimukseen.[7] Yleiset luonnolliset ja luovuttamattomat ihmisarvot ja ihmisen oman tahdon kunnioitus ohjaavat siis nykyään tutkimusetiikkaa. Yleinen periaate on se, että ihmisen tulee antaa tietoon perustuva suostumuksensa tutkimukseen. Suostumuksen ja tiedon tulisi ideaalisti kattaa kaikki tutkimuksen kuluessa tehdyt toimenpiteet ja prosessit, joita ovat muun muassa aineistojen kerääminen, käsittely, säilytys ja julkaiseminen. Tutkija voi huomata tässä – kuten joskus muuallakin tutkimuksessaan – normiristiriidan. Toisaalta alalla ehkä kehotetaan säilyttämään aineistot jatkotutkimuskäyttöön, mutta samalla halutaan lisätä tutkimukseen osallistuneiden henkilöiden oikeutta hallita omia tietojaan.

Suhteessa moneen muuhun ihmistieteen alaan kielitiede on ollut varsin turvallista osallistujille, ja tieteenala on säästynyt alaa ravisuttavilta

skandaaleilta. Ihmisarvon kunnioittamisen normi on kuitenkin johtanut myös kielentutkimuksessa arvioimaan esimerkiksi termejä, joilla kenttätutkimuksessa nimitetään tutkimukseen osallistuneita. Vaux ja Cooper (1999) sekä Munro (1998; 2000) toteavat, että esimerkiksi termillä *informant* (suom. *informantti*) voi olla englannin kielessä ikävä konnotaatio, sillä termiä käytetään myös poliisin informantista ja vakoiluinformantista. Tämä käyttöyhteys voi luoda vaikutelman ihmisestä, joka salaa, epäeettisesti tai vastoin kieliyhteisönsä tahtoa antaa tutkijalle informaatiota kieliyhteisön kielitiedosta. Toisaalta termi on yleisessä käytössä muun muassa sosiolingvistiikassa, eikä siihen ainakaan suomalaisessa kontekstissa liity väärinymmärryksen vaaraa. Tutkimukseen osallistuvista kielen puhujista on lisäksi käytetty mm. nimityksiä *kielenopas* ja *kielimestari*.

Toinen kielentutkimuksessa esillä ollut teema on tutkimukseen osallistuneiden ihmisten oikeus saada tunnustusta työstään ja tieteellisestä panostuksestaan. Vaux (2003) esittää, että kenttätutkimuksessa omasta tahdostaan ja omalla ajallaan toiminut henkilö voi kokea loukkaavana ja väheksyvänä sen, jos tutkija ei mainitsisikaan oleellista työtoveriaan lopullisessa julkaisussa. Vaikka tieteellinen raportti on useimmiten yksin kielentutkijan kirjoittama, on itse tutkimusprosessi ja aineisto yhteistyön tulos. Kenttätutkimuksessa kielenoppaan rooli on siinä määrin aktiivinen, aikaa vievä ja tutkimusta ohjaava, että Vauxin (2003) mukaan voidaan toisinaan puhua jopa kanssatutkijan roolista. Kielenopas voi toimia jopa vuosikausia kielentutkijan rinnalla tai opettajana. Kanssatutkijaa olisi alentavaa kohdella nimettömänä lopullisessa julkaisussa. Tämä Vauxin esittämä ajatus on kuitenkin normiristiriidassa sen tutkimuseettisen normin kanssa, joka esittää, että tunnistetietoja ei tule kerätä, säilyttää tai julkistaa tarpeettomasti. Tunnisteellisten aineistojen käsittelystä säädetään Suomessa henkilötietolaissa (523/1999). Lain 3. §:n mukaan henkilötiedolla tarkoitetaan ”kaikenlaisia luonnollista henkilöä taikka hänen ominaisuuksiaan tai elinolosuhteitaan kuvaavia merkintöjä, jotka voidaan tunnistaa häntä tai hänen perhettään tai hänen kanssaan yhteisessä taloudessa eläviä koskeviksi”. TENKin (2009)[8] mukaan perusperiaate tunnistetietojen keräämiselle ja säilyttämiselle on niiden tarpeellisuus tutkimuksessa.

Tunnustuksen antaminen kielenoppaan työlle sekä toisaalta henkilötietojen suojelu vaatii tasapainottelua. Joskus tieteessä tällaisessa on epäonnistuttu totaalisesti. Esimerkiksi niin sanottu Genie-tapaus osoitti, että tutkimuksen kuluessa on otettava huomioon tutkimukseen osallistuvien luonnolliset perustarpeet ja lasten erityinen asema. Genie on pseudonyymi kielettömälle tytölle, joka löydettiin vuonna 1970 Kaliforniasta. Tyttö oli ollut 20 kuukauden ikäisestä isänsä lukitsemana kellarissa. Päivisin lapsi oli sidottu pottaan ja öisin eräänlaiseen makuupussiin. Lapsi oli kasvanut 13 vuoden ikään ilman sosiaalisia kontakteja ja kuulematta puhetta. Lapsen löydyttyä myös tutkijat kiinnostuivat hänestä, ja häntä tutkittiin useissa eri hankkeissa (esim. Curtiss 1977). Rymer (1992) ja Jones (1995) pitävät Genie-tapausta esimerkkinä häikäilemättömästä tieteellisestä hyväksikäytöstä. Rahoituskauden aikana useat eri tutkijat altistivat tytön ankaraan testaukseen, seurantatutkimuksiin ja fyysisiin mittauksiin. Rahoituksen päätyttyä tutkijat hylkäsivät lapsen kasvattikotikierteeseen.

Erityisryhmiä, kuten lapsia, sairaita ja vähemmistöjä, koskevat niin muissa tieteissä kuin kielitieteessä nykyään omat tiukat tutkimusnorminsa, joiden päämääränä on muun muassa estää herkässä kasvuiässä olevien vahingoittaminen. Suomessa painava puheenvuoro lasten ja nuorten tutkimuksen etiikassa on Lagströmin, Pösön, Rutasen ja Vehkalahden (2015) toimittama artikkelikokoelma. Tutkimuseettisiä normistoja, joiden tavoite on kattaa erityistilanteet ja erityisryhmät, on kehitetty etenkin viimeisten vuosikymmenien aikana.

Kielitieteessä ylhäältä alas sovellettavia makroeettisiä normistoja tuotetaan nykyään kiihtyvällä vauhdilla yhdistyksissä, tiedelehdissä ja yliopistoissa. Moni näkee tässä myös huolestuttavan sääntöjentehtailutrendin. Pirttilän (2008, 85) mukaan eettiset toimikunnat ovat muuttuneet neuvoa antavista elimistä tutkimusta määrääviksi normittaviksi elimiksi, joiden ote ajan myötä vain tiukkenee. Mäkelän (2005; 2007) mukaan säädökset harvoin parantavat tutkittavien tosiasiallista asemaa. Tieteellisen tutkimuksen toteuttamista ne sitä vastoin hankaloittavat. Metodologisten normien ja eettisten säädösten välille onkin Mäkelän ja Pirttilän mukaan muodostumassa normikonflikti: tutkimuksen laatu kärsii, kun tutkimukseen sovelletaan väkisin siihen soveltumattomia normeja ja käsitteitä.

Esimerkiksi Kimmel (1996, 283) raportoi fonetiikan alaan kuuluvasta tutkijasta, joka ei saanut eettiseltä komitealta hyväksyntää eri korkeudella olevien äänten (joista yksikään ei ollut normaalia puheääntä korkeammalla eikä kuuloa vaurioittavaa) esittämiselle ihmishenkilöille. Peruste oli se, että tehtävä muodostaa riskin osallistujien hyvinvoinnille. Riski liittyi tutkimukseen osallistuvien henkilöiden mahdolliseen kyllästymiseen. Foneettinen tutkimus jouduttiin toteuttamaan eläimillä.

Linguistic Society of America -yhdistys (LSA) on varoittanut, että väkinäinen biolääketieteellisen tutkimuseettisen koodiston soveltaminen kielitieteeseen johtaa tutkimusresurssien tuhlaukseen ja heikentää tutkimuksen laatua. Suomessa Hallamaa, Launis, Lötjönen ja Sorvali (2006) ovat huomanneet, että eettisiä normistoja on yleensä laajennettu lääketieteellisestä tutkimuksesta muihin tutkimusaloihin usein kuitenkaan tutkimatta empiirisesti, esiintyykö näissä muissa tutkimustyypeissä ylipäätään samankaltaisia eettisiä huolenaiheita. LSA[9] on lisäksi toistuvasti todennut, että biolääketieteestä johdettu koehenkilörooli ja *henkilövahinko*-käsite ovat harvoin sovellettavissa kielitieteelliseen tutkimukseen. Yhdistyksen mukaan kielitieteessä tutkimukseen osallistuva ihminen on usein luonnollisessa tutkimusympäristössä, minkä vuoksi positivistinen eksperimentaatio-käsitteistö ei ole sovellettavissa. Lääketieteen kokeellisen tutkimusasetelman mahdollisuus aiheuttaa vahinkoa ihmiselle on luonteeltaan hyvin erilainen kuin esimerkiksi lingvistisen kenttätutkijan mahdollisuus tuottaa vahinkoa kielenoppaalle. On mielekästä erottaa manipuloitu ja objektivoitu koehenkilö (engl. *subject*) kyselytutkimukseen osallistuvasta henkilöstä (engl. *participant*). Mitä avoimempi on kyselytekniikka, sitä suurempi itsemääräämisvara vastaajalla on (Fossey, Harvey, McDermott & Davidson 2002). Kyselyyn osallistuva henkilö on siten usein tutkimuksessa subjektina, jonka ”ääni” tosiasiallisesti kuuluu (Konstenius 2014).

Kielentutkijan kannattaa siis pohtia erilaisten normien mielekkyyttä tutkimuksessaan: Turvaavatko ne todella osallistujien oikeuksia? Heikentävätkö ne tutkimuksen laatua? Myös hyvin vakiintuneet käsitykset tutkimusnormeista voivat yllättäen jossakin kontekstissa osoittautua epärationaalisiksi ja epäeettisiksi. Tutkimusetiikka elää murrosta, ja siinä tapahtuu nyt muutoksia.

Viimeisen kahden vuosikymmenen aikana internet on muuttanut sitä, kuinka tutkimusetiikka tänään ymmärretään. Kun ennen kieli jaettiin puhuttuun ja kirjoitettuun kieleen, jota puhutaan tai kirjoitetaan yksityisissä tai julkisissa yhteyksissä, tänään on erityisesti otettava huomioon se, millä tavalla kieli on medioitunut. Kielen sähköinen välitystapa ja se teknologia, joka välittää kieltä, vaikuttavat oleellisesti siihen, miten kieltä voi tai saa tutkia. Teknologioiden kehittyminen on myös mahdollistanut aivan uusia tapoja tutkia kieltä. Aineistojen kerääminen, käsittely ja muokkaaminen digitaalisissa ympäristöissä herättävät siksi uudenlaisia eettisiä huolenaiheita, joita ei aikaisempien tutkimusten perusteella ole voitu ennakoida. Hyvin uudenlainen keskustelu ihmisen oikeuksista ja niiden kunnioittamisesta on meneillään digitaalisissa tieteissä.

## 4.5. Kielentutkimus digitaalisessa ympäristössä

Jo ennen internetiä tieteellinen tutkimus siirtyi teknologiseen ympäristöön. Tutkijoiden kammioissa kirjoituskoneet vaihtuivat tietokoneisiin, arkistot levykkeisiin ja metalliset avaimet salasanoihin. Tutkijan materiaaliset työkalut vaihtuivat sähköisiin miltei huomaamatta. Sähköisten työkalujen yleistyminen mahdollisti internetin kehittymisen, joka vuorostaan muutti lähtemättömästi ihmisten tapaa kommunikoida keskenään. Se sai aikaan ajassamme merkittävän kulttuurisen murroksen. Yhteiskunta muuttui informaatioyhteiskunnaksi, ja uudet teknologiat tulivat osaksi lähes kaikkea inhimillistä toimintaa. On vaikea enää löytää tietoyhteiskunnan jäsenen elämässä aluetta, jota teknologia ei edes välillisesti koskettaisi.

Tietoyhteiskunta ei perustu yhteen teknologiseen oivallukseen. Internetin erityispiirre on juuri se, että se ei ole yksi teknologia vaan alati muuttuva teknologinen myriadi. Sitä ei voi siksi suoraan verrata painokoneeseen tai höyryjunaan. Yhden mekanismin ymmärtäminen ja pohtiminen ei ratkaise uusien teknologioiden myötä esiintyviä eettisiä huolenaiheita. Uusien teknologioiden erityispiirre on se, että ne mahdollistavat yhä uusia teknologisia innovaatioita kiihtyvällä vauhdilla.

Moorin (1985) mukaan uusiin teknologioihin kuuluu, että uusia piirteitä ei voi ennakoida vanhojen perusteella, ja siksi vanhat lait ja normit eivät automaattisesti kata uusien teknologioiden herättämiä eettisiä huolenaiheita, vaan ne täytyy tutkia erikseen. Buchananin (1999) mukaan näyttää siltä, että tässä uudessa digitaalisessa tieteellisessä ympäristössä staattisten ja helposti vakiinnutettavien tutkimuseettisten koodistojen kehittäminen on nyt epärealistista.

Teknologioiden nopeassa kehityksessä makroeettiset normistot eli ylhäältä alas, kaikkiin tutkimusasetelmiin yhtäläisesti ja kategorisesti sovellettavat normit alkavat näyttää menneiden aikojen hankkeilta. Digitaalisesta diversiteetistä on tullut uusi todellisuus, mikä tarkoittaa sitä, että on yhä pontevammin tiedostettava mikroetiikka. On huomioitava jokaisessa tutkimuksessa erikseen ja eksplisiittisesti tieteellisissä julkaisuissa, miten tutkimuksen kuluessa käytetyt teknologiat ja tekniikat otetaan huomioon (Kubanyiova 2008). Tutkijan on perusteltava ja yksilöitävä tutkimuseettiset ratkaisut, joilla hän takaa sen, että tutkimus on toteutettu eettisesti. Mikroetiikka, tutkimusjulkaisuihin integroitu moraalifilosofinen pohdinta, näyttäisi olevan erityisesti digitaalisen tieteen ominaisuus.

Mikroetiikassa oleellista on tutkimuksen kuluessa vastata joukkoon tutkimuseettisiä kysymyksiä. Tutkijalla on hyvä olla kyky tulla tietoiseksi tutkimuseettisistä huolenaiheista ja adaptoida tutkimusasetelma eettisten päätelmiensä mukaiseksi. Ulkoapäin tulleista normeista tutkijan täytyy digitaalisessa ympäristössä siirtyä eksplisiittiseen tutkimuseettiseen argumentaatioon. Tutkimuksen kuluessa tutkimusasetelmaa voi joutua muuttamaan uusien eettisten huolenaiheiden ilmetessä. Bruckman (2012) ennakoi myös muuttuvaa suhdetta eettisten lautakuntien ja tutkijoiden välillä: tutkijoiden on aktiivisesti ylläpidettävä eettisten lautakuntien tietoja uusista eettisistä huolenaiheista muuttuvissa sähköisissä ympäristöissä. Eilisen valvovasta isoveljestä on tulossa oppilas.

Sähköisten ympäristöjen tutkimusetiikkaan liittyy useita erityiskysymyksiä, vaikka oletettavasti ihmisen luonnolliset oikeudet ovatkin pysyneet samoina. Ihmisellä pitäisi edelleen olla oikeus valita, osallistuuko hän tutkimukseen. Hänellä on oikeus yksityisyyteen, autonomiaan, henkilöönsä liittyviin tietoihin, turvallisuuteen ja hyvinvointiin.

Oikeuksien kunnioittaminen ei kuitenkaan ole teknologiavälitteisissä ympäristöissä yksioikoista, sillä suuri(n) osa internetin liiketoimintamalleista perustuu käyttäjien datan keräämiseen, analyysiin ja myyntiin: käyttäjä maksaa tietyn "hinnan", jonka vastikkeeksi hän saa ilmaiseksi käyttöönsä teknologiaa. Onko käyttäjä tällöin luopunut oikeuksistaan yksityisyyteen, autonomiaan ja digitaaliseen henkilöönsä? Jos on, mitä tutkimusetiikassa tarkalleen ottaen silloin suojellaan? Usein internetin käyttäjien käytöstä seurataan heidän tietämättään ja dataa kerätään kertomatta, mitä tietoja on kerätty, missä niitä säilytetään ja mihin tarkoitukseen niitä käytetään. Erään käsityksen mukaan internet onkin alusta alkaen ollut eräänlainen käyttäytymistieteellinen laboratorio, jossa kaupalliset yritykset ovat voineet vapaasti testata ihmisillä erilaisten ilmiöiden vaikutuksia (Ross 2014; Hunter & Evans 2016).

Juuri internetissä laajassa skaalassa tehdyt kokeet ja kuluttajien suostumus datojensa keräämiseen ja käyttöön ovat saaneet monet tieteenalojen yhdistykset pohtimaan tiukkoja tutkimuseettisiä normejaan. Esimerkiksi American Psychological Association (APA 2010)[10], joka on perinteisesti ollut erityisen tiukka ihmiseen kohdistuvan tutkimuksen normeista, on jossain määrin joustanut periaatteistaan online-tutkimuksissa. Esimerkiksi vuonna 2012 keskusteltiin laajasti siitä, että tutkijat olivat manipuloineet ihmisten tunnereaktioita Facebookissa. Laajamittainen tunnetutkimus oli toteutettu 600 000 henkilöllä, ja sen kautta saatiin pätevää empiiristä tietoa siitä, miten tunnekokemukset vaikuttavat ja leviävät populaation tasolla. Tätä *Massive-scale Emotional Contagion Through Social Networks* -tutkimusta (PNAS 2014) puolustettiin APA:ssakin toisaalta sillä, että tutkimusta ei olisi voitu toteuttaa millään muulla tavoin, ja se oli siksi tarpeellinen. (Ross 2014.) Toisaalta tutkimusta puolustettiin myös sillä, että kaupalliset yritykset tekevät päivittäin tuhansia samanlaisia tutkimuksia kuluttajilla, ja akateemisillakin tutkijoilla tuli olla samat lähtökohdat.

Tämä akateeminen tutkimus nosti suuren yleisön tietoisuuteen sen seikan, että kaupalliset yritykset todella tekevät päivittäin tuhansia eksperimenttejä suurilla populaatioilla omilla verkkosivustoillaan. Akateemiset tutkijat usein tekevät omia tutkimuksiaan pienemmällä skaalalla, julkisesti ja vähemmän manipuloiden. Tutkimusetiikassa on vastattava

siihen, miksi hyväksytään yksi tutkimustyyppi rahoituksen perusteella mutta ei toisenlaista. Niiniluoto kutsuu *saleksi* tutkimusta, jota tehdään suuren yleisön tietoisuuden ulottumattomissa omaa voiton pyyntiä tavoitellen (ks. Johdannon luku 2.2. tk.). Miksi hyväksyttäisiin sale-tutkimukset internetissä mutta ei tieteellisiä, julkisella rahoituksella suoritettuja ja yhteishyvään pyrkiviä tutkimuksia?

Kaikki käy -asennetta internetissä ovat puolustaneet erityisesti libertarianistiset yhdistykset. Tämä on erityisesti Yhdysvalloissa suosittu poliittinen näkökanta, jonka mukaan valtion ei tulisi rajoittaa kansalaisten tai yritysten pyrkimyksiä. Esimerkiksi Electronic Frontiers Foundations (EFF) on lobannut internetin toimijoille ”vapautta” kansallisvaltioiden lainsäädännöstä. Kansalaisen vapauksien nimissä – tai sijaan – puolustetaankin yritysten vapautta toimia ilman valvontaa ja sääntelyä internetissä. Nykypäivänä juuri yritysten häikäilemätön ja epäeettinen toiminta usein rajoittaa yksilön vapauksia. Vaihtoehtona onkin esittää, että yritykset noudattaisivat tutkimuksissaan internetissä tieteen tutkimusetiikkaa.

Kaikille yhteisten pelisääntöjen puuttuminen on akuutti ongelma, jonka ratkaisemiseen myös kielitieteellisen yhteisön tulisi osallistua. Kielentutkimuksella voisi ajatella olevan erityinen ääni digitaalisen tieteen etiikassa. Juuri kielentutkijoilla on pätevää tietoa siitä, millä tavoin erilaiset kielelliset ilmiöt ovat sidoksissa ihmiseen ja tämän erilaisiin elämänalueisiin eli siihen, miten kieli vaikuttaa yhteiskunnassa, yhteisöissä ja vuorovaikutuksessa. Kielitieteen näkökulmasta tutkimusetiikassa nyt esitetty kategorinen kysymys siitä, onko sähköinen tutkimus tekstintutkimusta vai ihmistutkimusta, on ehkä liian suppea. Kielitieteen näkemys kielen ja ihmisen suhteesta on tällaista joko–tai-kysymystä laajempi.

Wolframin (1993) mukaan kielentutkijalla on vastuu kieliyhteisöä kohtaan myös sähköisissä ympäristöissä. Jos tutkija huomaa, että yhteisöä tutkitaan tai kohdellaan epäeettisesti, tulisi hänen toimia yhteisön eduksi ja vaatia sellaista yhteiskunnallista sääntelyä, jonka avulla yhteisön etuja vaalitaan. Jos tutkija huomaisi esimerkiksi, että jotakin alkuperäiskansaa altistettaisiin rasittavalle, invasiiviselle (henkilön itsemääräämisoikeuteen kajoavalle) ja epäkunnioittavalle tutkimukselle, olisi hänen velvollisuutenaan toimia siten, että löydettäisiin sellaiset

yhteiset pelisäännöt, joiden mukaan kaikki tutkijat toimisivat ja jotka takaisivat kieliyhteisön hyvinvoinnin. Sama periaate pätee sähköisissä ympäristöissä.

Wolframin (1993) periaatteen noudattaminen on paljastanut tutkimusetiikkaan toisinaan implisiittisesti piiloutuneen holhoavan asenteen: tutkijat uskovat tietävänsä, mikä on tutkittaville parasta. Internetin toimijoilla on yleensä oma ääni. Kun tutkijat ovat ehdottaneet omasta mielestään eettisiä toimintakoodeja sähköisiin ympäristöihin ja niissä toimiviin yhteisöihin, ne harvoin ovat olleet käyttäjien oman tahdon ja tavoitteiden mukaisia (Zimmer 2010; Zimmer & Proferes 2014). Holhoavaa asennetta tutkimukseen osallistuviin henkilöihin ei sitäkään voi pitää täysin eettisenä.

Digitaalisuutta koskeva tiede on aiheuttanut sen, että tutkimusetiikka elää murrosvaihetta. Seuraavaksi esille nostetaan kolme kysymystä, joihin digitaalisessa mikroeettisessä tutkimuksessa yleensä vastataan: 1) onko teknologia työkalu vai virtuaalinen tila tutkimuksessa, 2) onko tutkimuksella suora vai epäsuora suhde ihmisiin ja 3) mitä on tunnisteellinen ja ei-tunnisteellinen data. Kysymykset eivät ole tyhjentäviä; teknologian kehittyessä ilmaantuu uusia kysymyksiä. Näihin kysymyksiin ei myöskään ole yhtä oikeaa vastausta. Oleellista on kysymysten eettisen luonteen tiedostaminen ja niiden asemointi tutkimusta ohjaavassa moraalifilosofisessa päättelyssä.

### 4.5.1. TEKNOLOGIA: TYÖKALU VAI VIRTUAALINEN TILA TUTKIMUKSESSA?

Teknologia voidaan tutkimuksessa ymmärtää joko a) työkaluksi tai b) virtuaaliseksi tilaksi, ”paikaksi”. Seuraavassa tarkastellaan, millä tavoin nämä kaksi erilaista tapaa ymmärtää teknologia vaikuttavat eettiseen pohdintaan.

Kun uusia teknologioita käytetään **työkalun** tavoin, silloin yleensä ohjelmistoja käytetään informaation hakuun ja säilytykseen sekä datan lajitteluun, koodaamiseen, käsittelyyn ja analysointiin. Tällöin tutkijan on tiedostettava **teknologian etiikka**. Teknologian etiikka tarkoittaa, että teknologian rakenteissa on ennakko-oletuksia tai olettamia ihmisestä, ihmisryhmistä ja näiden oikeuksista. Esimerkiksi pankkiautomaatti

olettaa keskipituisen ihmisen. Sen rakenteista voidaan päätellä, ottaako se huomioon vähemmistöt, kuten lyhytkasvuiset, liikuntavammaiset tai sokeat. Samalla tavalla ohjelmistoissa on sisäänrakennettuja arvoja, jotka syrjivät tai suosivat eri väestöryhmiä. Esimerkiksi informaation hakuun käytetyt hakukoneet eivät ole neutraaleja. Googlen algoritmit suosivat sivustoja, jotka ovat helposti koneluettavia, toteuttavat Googlen omia layout-ohjeita ja joissa on AdWords-toiminto[11], mistä syystä Google on viimeisten vuosien aikana haastettu useasti EU-tuomioistuimeen. Google jättää pois kilpailijat ja verkkosivustot, jotka eivät maksa Googlelle hauista, tai asettaa ne tarkoituksellisesti huonoon asemaan (Introna & Nissenbaum 2000).

Digitaaliseen viestintään kohdistuvan kielitieteen tutkimusetiikassa tutkijoiden on tultava tietoisiksi käyttämiensä työkalujen arvorakenteista ja pyrittävä neutralisoimaan vääristymät. Mikroeettisessä lähestymistavassa tulisi argumentoida ja eksplikoida sitä, miten teknologiaan sisäänrakennetut arvovääristymät on otettu tutkimuksessa huomioon.

Kun uudet sähköiset teknologiat ymmärretään **virtuaaliseksi tilaksi**, yleensä viitataan sellaisiin teknologioihin kuin peleihin, virtuaalisiin maailmoihin, uutisryhmiin, blogeihin, mikroblogeihin, verkkokursseihin, foorumeihin, keskusteluapplikaatioihin yms. Bruckman (2002; 2006; 2012) kehottaa pohtimaan käytössä olevia tilaa ja tilallisuutta kuvaavia ilmauksia: mitä ne kertovat virtuaalisesta tilasta, ja mitä ne jättävät kertomatta? Sähköinen media vaikuttaa kommunikaation tapaan eri tavoin: se edellyttää reittejä ja asettaa läpipääsemättömiä seiniä, ovia ja avaimia. Esimerkiksi salaisiin[12] Facebook-ryhmiin, joihin voi liittyä vain kutsusta ja jotka ovat erilaisten teknologisten suojamuurien, pääsyvaatimusten ja salasanojen takana, liittyy suurempi oletus yksityisyydestä kuin tiloihin, joissa tällaisia eristäviä toimintoja ei ole. Suljettu tai salainen internetryhmä voi tuntua käyttäjistään ”koti-illalta ystävien kesken”, ja ryhmässä usein keskustellaan intiimisti olettaen, että sinne ei tunkeudu ulkopuolisia eikä siellä puhuttuja asioita levitellä muille.

Käyttäjien kokemukset yksityisyydestä ja tosiasiallinen yksityisyys voivat olla ristiriidassa. Hudson ja Bruckman (2005) suorittivat laajan internetkäyttäjätutkimuksen, jossa verrattiin käyttäjäkokemuksia yksityisyydestä ja todellista yksityisyyttä. He tulivat siihen tulokseen,

että näiden välillä oli yleensä suuri ristiriita ja että käyttäjillä oli vain rajallinen tietoisuus siitä, millä tavoin teknologinen alusta tunkeutui heidän yksityisyyteensä. Suljettujenkin ryhmien käyttäjien IP-osoitteet, logit ja muut tunnisteelliset tiedot ovat palveluntarjoajan käytettävissä. Useimmat suuret yritykset analysoivat käyttäjiensä käyttäytymistä – joskus jopa sivuilla käytyjä keskusteluja. Keskustelut ja niiden sisältö on useissa maissa laillisesti yrityksen omaisuutta, mikäli näin on esitetty käyttäjäehdoissa. Käyttäjän kokema yksityisyys, käyttäjäehdoissa luovutetut oikeudet yksityisyyteen ja varsinaiset yksityisyysasetukset voivat siten kaikki olla keskenään erilaiset. Tila, joka käyttäjilleen ilmenee ”koti-iltana ystävien kesken”, voikin datankäytön valossa pikemminkin näyttää tilalta, johon on asennettu kameroita ja äänityslaitteita ja jossa vuokranantaja tekee salaa merkintöjä vuokralaisen käyttäytymisestä vastaisuuden varalle. Digitaalisessa tutkimusetiikassa on usein yritetty suojata käyttäjien kuviteltua yksityisyyttä tai yksityisyyden kokemusta. Teknologian medioimien tilojen kompleksisuus tulisi kuitenkin kokonaisuudessaan ottaa huomioon moraalifilosofisessa päättelyssä. Kun suostumusta pyydetään tutkimuksen kohteilta, verkkosivun ylläpitäjältä tai eettiseltä komitealta, on mielekästä problematisoida näitä erilaisia tiloja ja nostaa ne keskustelun kohteeksi.

Sähköiset ympäristöt eroavat aikaisemmista medioista erityisesti juuri siinä, että aikapaikkaiset tekijät ovat muuttuneet ja aikapaikkaisuutta tilallisena kokemuksena luovat nyt yksityiset yritykset. Kommunikaatio tapahtuu yksityisten yritysten tiloissa, joihin yrityksillä on tekijänoikeus, eikä suurella yleisöllä ole mahdollisuutta tarkastella arkkitehtuurin tavoin toimivia ohjelmistoja.

Perinteiset tilailmaukset sopivat huonosti kuvaamaan uusia teknologian medioimia ympäristöjä. Samaan aikaan emme kuitenkaan vielä osaa tehdä moraalifilosofista päättelyä ilman niitä. Esimerkiksi yksityisyys on ihmisen oikeus, joka aikaisemmissa yhteiskuntamuodoissa on määritetty juuri tilan kautta. Aikapaikkaisessa offline-ympäristössä ihmisellä on oikeus vaatia, että kukaan ei kurki hänen kotinsa ikkunoista, mutta julkisessa kulkuvälineessä hänen täytyy suostua olemaan myös toisten katseen kohde. Tilametaforien vaikeasta suhteesta digitaalisen tieteen moraalifilosofiaan tulee siten olla tietoinen, ja metaforia tulisi

problematisoida tutkimuksen kuluessa. Sähköinen ympäristö ei ole suoraviivaisesti aikapaikkainen tila, joten sopii kysyä, miten ja millaisin ohjelmistoin tila on luotu. Mitä nämä ohjelmistot ihmisen käytöksestä tässä tilassa välttämättä lukevat, tietävät ja tallentavat?

### 4.5.2. ONKO TUTKIMUKSELLA SUORA VAI EPÄSUORA SUHDE IHMISIIN?

Ihmistä tai ihmisen toimintaa koskeva tutkimus voi olla eri astein kajoavaa eli invasiivista. Tällä tarkoitetaan sitä, että tutkimustoiminta saattaa tarkoituksellisesti tai tahattomasti tunkeutua ihmiseen, ihmisen elinpiiriin tai ihmisen harjoittamaan toimintaan. Käsitettä *invasiivinen* on tyypillisesti käytetty lääke- ja hoitotieteissä, mutta se soveltuu myös ihmistieteisiin ja humanistiseen tutkimukseen. Seuraavaksi tarkastellaan sitä, mikä on kielentutkimuksen vaikutus kielianeiston tuottaneisiin ihmisiin ja ihmisryhmiin erityisesti verkkoympäristössä.

**Läsnäoleva verkkotutkimus** (engl. *engaged web-based research*) on usein tutkimustyyppi, jossa traditionaaliset testit ja kyselyt on tuotu verkkoympäristöön. Sillä on yleensä välitön interaktiivinen ja invasiivinen suhde ihmisiin. Tähän tutkimustyyppiin myös usein sovelletaan traditionaalisia offline-tutkimuskoodeja. **Ei-invasiivinen verkkotutkimus** (engl. *non-intrusive web-based research*) jättää asiaintilat ennalleen vaikuttamatta niihin suoraan (Kitchin 2008). Erilaiset big data -työkalut (engl. *harvesting, mining, scraping*) ja internetkäyttäytymistä seuraavat teknologiat (engl. *profiling, observation, recording*) keräävät yleensä olemassa olevasta datamassasta piirteitä. Nämä teknologiat toimivat internetkäyttäjän havainnon ulkopuolella, huomaamattomasti, eikä henkilöllä ole interaktiivista suhdetta tutkimuksen kanssa. Tutkimustyyppi ymmärretään siksi epäsuoraksi ja ei-invasiiviseksi: tutkimus ei muuta asioiden kulkua reaalitodellisuudessa.

Jako invasiiviseen ja ei-invasiiviseen tutkimukseen on nähty jossain määrin samanlaisena kuin jako, joka on tehty ihmiseen kohdistuvan tutkimuksen ja tekstintutkimuksen välille (McKee & Porter 2009). Ihmistutkimuksessa tutkijalla on interaktiivinen suhde tutkittaviin, ja tekstintutkimuksessa tutkija analysoi ihmisten kirjallisia tuotoksia. Kategorioiden välille voi myös olettaa jatkumon, jossa tutkittavaa aineistoa

voi jäsentää edelleen sen mukaan, onko aineisto yksityistä vai julkista, sensitiivistä vai neutraalia, interaktion vai julkaisemisen tulosta jne.

Klassinen jako ihmiseen kohdistuvaan tutkimukseen ja tekstintutkimukseen on kuitenkin ongelmallinen sähköisissä ympäristöissä kahdesta syystä. Ensiksi ihmistä ja hänen verkkokäyttäytymistään voi seurata reaaliajassa ja pitkienkin ajanjaksojen yli, mikä ei ollut aikaisemmin mahdollista. Sähköisessä ympäristössä voi määrittää, mitä joku lukee, katsoo tai sanoo, kuinka kauan ja kenen kanssa sekä mikä hänet saa kiinnostumaan, poistumaan tai jäämään sivulle. Yrityksillä on yksityiskohtaista tietoa käyttäjistään tavalla, joka ei aikaisemmin ollut mahdollista. Tämä tarkkailu ei ole invasiivista, eikä verkkosurffaaja yleensä edes tiedä, kuinka monta ohjelmaa häntä seuraa tai mitä tietoja ne keräävät hänestä, hänen ystävistään ja perheestään.

Toisekseen ihminen ja hänen ilmaisunsa ovat periaatteessa toisistaan erillisiä. Sähköisessä ympäristössä ihmisen tuotokset on aina medioitu teknologian kautta, ja ne ovat siten hänestä irrallaan. Tuotokset myös usein jäävät elämään digitaalista jatkoelämää niiden yritysten ja tutkijoiden arkistoissa, joihin tuotokset on tallennettu. Sähköiseen ympäristöön jää digitaalinen jälki tavalla, joka on historiassa ainutlaatuinen. Tavallisen ihmisen arkielämästä tallentuu jälkiä lähes automaattisesti tuleville sukupolville.

White (2002) on sitä mieltä, että kielelliset ilmaukset sähköisissä ympäristöissä ovat tekstejä, ja vastustaa sitä, että online-tutkimuksissa käsiteltäisiin tekstejä aivan kuin ne olisivat fyysisiä henkilöitä eikä erotettaisi elävää subjektia ja kulttuurisia tuotoksia toisistaan. Hänen mukaansa digitaalinen tutkimus ei edes periaatteessa voi olla ihmiseen kohdistuvaa tutkimusta, sillä internetissä ei ole fyysisiä henkilöitä.

Hyvin toisenlaisen näkemyksen tarjoaa Fairfield (2012), jonka mukaan *henkilö*-käsitettä olisi internetissä laajennettava, ei suinkaan kavennettava. Hän on käyttänyt termiä *human harming research* sellaisesta tutkimuksesta, joka on epäsuoraa mutta joka voi aiheuttaa vahinkoa tai tuskaa ihmiselle. Esimerkiksi virtuaalisissa maailmoissa avattaren (profiilikuva) vahingoittaminen voidaan tulkita ihmisen identiteettiä loukkaavaksi teoksi. Vaikka avatar ei ole ihminen vaan representaatio jostakin ihmiselle itselleen tärkeästä asiasta (Fairfield 2012), se on

kuitenkin osa hänen henkilöään. Fairfield (mt.) laajentaisi *henkilön* käsitettä koskemaan virtuaalisissa ja sähköisissä ympäristöissä sellaisia ihmiselle tärkeitä asioita, joita vahingoittamalla ihmiselle voi tuottaa tuskaa. Ihmisen identiteetti rakentuu sähköisissä ympäristöissä. Fairfield (mt.) ehdottaa, että esimerkiksi nimimerkille tai henkilökohtaiselle verkkosivustolle aiheutettu vahinko vertautuisi henkilövahinkoon.

On myös esitetty käsite *human non-subjects research* (*subjektiton ihmistiede*). Brothers ja Clayton (2010) ovat käyttäneet termiä tutkimuksesta, jossa ihmistä koskevasta datasta on poistettu kaikki tunnisteet ja data käsitellään ja analysoidaan ihmisestä riippumatta. Tällöin data yhä selkeästi koskee ihmistä, mutta siinä ei enää esiinny persoonallisia subjekteja, tunnistettavia henkilöitä. Datamassaa analysoidaan kokonaisuutena. Tämä subjektiton ihmiskorpus toimisi heidän näkemyksensä mukaan hieman samoin kuin kielitieteessä korpustutkimus: sen käsittely ei tuota harmia tai haittaa niille yksittäisille ihmisille, jotka tekstin ovat tuottaneet.

Bruckman (2014) on kehottanut tutkijoita aktiivisesti osallistumaan tutkimuskohteen määrittelyyn nykyisessä tutkimuseettisessä murroksessa ja kehittämään myös uusia käsitteitä ja kategorioita. Vaarana on, että jos vain harvat osallistuvat keskusteluun, käy niin kuin biolääketieteessä: yhdestä tutkimustyypistä laajennetaan muihin tutkimuksiin näihin soveltumattomia toimintakoodeja ja heikennetään tieteellisen tutkimuksen laatua ja toimintaedellytyksiä. Digitaalisessa ihmistieteessä tutkijan täytyy uudella käsitteellisellä välineistöllä osallistua digitaalisen ympäristön tilallisen luonteen kuvaukseen, tulkintaan ja määrittämiseen.

### 4.5.3. MITÄ ON TUNNISTEELLINEN JA EI-TUNNISTEELLINEN DATA?

Tutkimusetiikassa periaatteena on pidetty, että tutkimus tulee suorittaa siten, että tutkimukseen osallistuneita henkilöitä ei voida tunnistaa ja ettei kerätä tai julkaista sellaisia tietoja, jotka saattaisivat heidät vaaraan, tahraisivat heidän mainettaan, hankaloittaisivat työllistymistä tai saattaisivat juridiseen edesvastuuseen. Tunnisteellisia tietoja saa kerätä vain erityisen tarpeellisissa tapauksissa, ja silloinkin näihin tietoihin

kohdistuvat ankarat salausvaatimukset. Suoria tunnisteita ovat nimi, henkilötunnus, henkilönimen mukainen sähköpostiosoite ja biometriset tunnisteet, esimerkiksi sormenjälki tai kasvokuva. Epäsuoriksi tunnisteiksi nimitetään sellaisia tietoja, jotka yhdistettynä muihin tietoihin mahdollistavat henkilön tunnistamisen. Tällaisia ovat esimerkiksi sukupuoli, ikä, asuinpaikkakunta tai ammatti.

Sähköiset ympäristöt ovat muuttaneet *tunniste*-käsitteiden sisältöä. Sähköisillä päätteillä, internetyhteyksillä ja -käyttäjillä on teknologisissa ympäristöissä sellaisia yksilöllisiä tunnisteita, joiden perusteella sähköinen kommunikaatio on ylipäätään mahdollista. Laki ei toistaiseksi kiellä näiden tunnisteiden keräämistä, mutta sähköisissä ympäristöissä ne ovat viime kädessä sosiaaliturvatunnukseen verrattavissa olevia yksilöiviä tunnisteita.

Epäsuorat tunnisteet ovat nekin kategoriana laajentuneet. Tällä hetkellä on epäselvää, mikä ei ole epäsuoraa tunnistetta sähköisessä ympäristössä. Sähköiset ympäristöt ovat kehittyneet monikanava-alustoiksi, joissa tietoja linkitetään. Haku pöytäkoneelta linkittyy matkapuhelimen paikkatietoihin, ja tabletissa avattu sähköposti vaikuttaa matkapuhelimen uutissyötteeseen. Ihmisen musiikkimaku voi kertoa hänen seksuaalisesta suuntautumisestaan, uutisvalinnat asuinpaikasta ja automerkki poliittisesta kannasta.[13] Sellainen tieto, joka eilen näytti neutraalilta, onkin tänään sellaista, jonka perusteella henkilö voidaan identifioida. Ihminen voidaan nyt tunnistaa uudenlaisen datan perusteella ja hänestä tehdään sellaisia päätelmiä, joita hän ei ehkä haluaisi itsestään julkistaa. Miten ennustaa huomisen teknologisia innovaatioita ja niiden vaikutuksia ihmisiin, dataan ja tutkimuksen etiikkaan?

## VIITTEET

1 Facebookia vuonna 2018 ravistellut Cambridge Analytica -skandaali on vain yksi esimerkki monien joukossa siitä, millaisia seurauksia voi olla sillä, että liikeyrityksille suodaan sellaisia vapauksia, joita ei yhteiseen hyvään pyrkivälle akateemiselle tutkimukselle suotaisi. Cambridge Analytica käytti miljoonien Facebook-käyttäjien yksityistä informaatiota näiden henkilöiden ja heidän ystäviensä persoonallisuuspiirteiden profiloimiseen ja manipuloimiseen siten, että heihin saattoi kohdistaa yksilöllistä tunteisiin vaikuttavaa poliittista mainontaa ja siten vaikuttaa kansallisella tasolla eri maiden demokraattisiin vaaleihin.

2 On hyvä huomata, että insinöörialoilla on ollut käytössä termi *mikroetiikka* samassa mielessä kuin tässä tekstissä käytetään *tutkimusetiikkaa* ja vastaavasti termi *makroetiikka* samassa merkityksessä kuin tässä käytetään *laajan vastuun etiikkaa* (ks. esim. Hudspith 1991).

3 https://www.tenk.fi/.

4 Tierneyn (2000) mukaan antropologian alan ja kielitieteen "epäpoliittiset" asenteet kenttätyössä ovat johtaneet ihmisoikeusloukkauksiin. Esimerkiksi Yanomameille (alkuperäiskansa Amazonian sademetsäalueella Brasilian ja Venezuelan rajalla) ei tarjottu antropologien ja lingvistien mukanaan tuomiin kulkutauteihin rokotteita sairauksien immunisoimiseen tai lääkkeitä niiden hoitoon. Tierney (mt.) esittää, että tutkija ei voi teeskennellä olevansa ulkopuolinen neutraali tarkkailija, sillä hänen läsnäolostaan on välttämättä aina seurauksia.

5 LSA: https://www.linguisticsociety.org/sites/default/files/Ethics_Statement.pdf.

6 Ihmisarvoa loukkaavia lääketieteellisiä kokeita tehtiin laajasti myös keskitysleirien ulkopuolella sairaaloissa ja orpokodeissa – niin natsi-Saksassa kuin muuallakin maailmassa. Esimerkiksi Heidelbergin psykiatrisessa sairaalassa Carl Schneider tappoi lapset kokeiden jälkeen tutkiakseen histopatologisia diagnoosejaan. Wienissä Elmar Turk ja Heinrich Gross käyttivät lastensairaaloissa samaa menetelmää diagnoosiensa varmistamiseksi. Vuosina 1942 ja 1943 Itävallan lastensairaaloissa tapettiin lääketieteellisissä kokeissa ainakin 800 lasta (Weindling 2008). Keskitysleirien ulkopuolella tehdyistä kokeista vain harva päätyi oikeuteen. Epäeettisiä kokeita ja niiden tutkimustuloksia on kyseenalaistettu vain vähän. Carl Schneiderin miniä perusti 1946 väitöskirjansa appensa tutkimustuloksiin. Heinrich Gross julkaisi teoriansa vuonna 1952, ja Elmar Turkin paloittelemien lasten ruumiinosia käytettiin eri yliopistoissa (Weindling 2008).

7 YK:n ihmisoikeusjulistuksen artikla 7, *International Covenant on Civil and Political Rights*, lausuu seuraavasti: "No one shall be subjected to torture or to cruel, inhuman or degrading treatment or punishment. In particular, no one shall be subjected without his free consent to medical or scientific experimentation."

8 https://www.tenk.fi/sites/tenk.fi/files/eettisetperiaatteet.pdf.

9 Esim. https://www.linguisticsociety.org/sites/default/files/lsa-stmt-research-human-subjects.pdf.

10 APA: www.apa.org.

11 Googlen algoritmit lukevat sivustojen sisältöä, mistä syystä ne antavat myös ohjeita, millaiset sivustot ovat helposti koneluettavia. Yritysten ja sivustojen tarve tulla nähdyksi

Googlen hauissa on johtanut sivustojen hyvin homogeeniseen ulkoasuun. Tietynlaisen ulkoasun omaavaan sivustoon voidaan koneellisesti sijoittaa mainoksia. AdWords on tällaiseen soveltuva mainosohjelma, joka sijoittaa sivuille käyttäjäprofiloinnin mukaista kaupallista mainossisältöä. Näistä mainoksista Google saa pääasiallisen liikevoittonsa.

12 Facebook-ryhmiä on vuonna 2019 kolmessa kategoriassa. ”Avoin” tarkoittaa, että ryhmä, ryhmän jäsenet ja kommentit näkyvät kaikille, myös ei-jäsenille. ”Suljettu” tarkoittaa, että ryhmä ja sen jäsenet näkyvät yleisölle mutta keskustelu näkyy vain jäsenille. ”Salainen” tarkoittaa, että ryhmän jäseneksi voi tulla vain kutsusta eikä ryhmää, sen jäseniä tai keskustelua näy ulkopuolisille. Monilla muilla foorumeilla on yleensä vain kaksi vaihtoehtoa: avoin keskustelu, joka on julkinen, ja suljettu keskustelu, joka näkyy vain jäsenille. Hudson ja Bruckman (2005) käyttävät termejä *avoin* ja *suljettu* tässä dikotomisessa merkityksessä.

13 Juuri tällaisia epäsuoria tunnisteita käytettiin Cambridge Analytica -tapauksessa vuonna 2016 USA:n presidentinvaalien yhteydessä identifioimaan ihmisten persoonallisuustyyppejä. Epäsuorat tunnisteet ennustivat luotettavasti sitä, miten ihminen reagoi erilaisiin kielellisiin ja visuaalisiin viesteihin. (Cadwalladr & Graham-Harrison 2018.)

## KIRJALLISUUS


AAAL. 2011. AAAL resolution against discrimination on the basis of accented speech. [verkkoaineisto]. [viitattu 10.4.2019]. Saatavissa: https://www.aaal.org/position-statements#Feb-201.

ALAA. 1998. The Applied Linguistics Association of Australia: Recommendations on good practice in Applied Linguistics. [verkkoaineisto]. [viitattu 11.4.2019].

APA = American Psychological Association. 2010. *Publication Manual of the American Psychological Association*. (6. p.). Washington (DC): American Psychological Association.

BAAL = The British Association for Applied Linguistics. 2006. *Recommendations on good practice in Applied Linguistics*. [verkkoaineisto]. [viitattu 11.4.2019]. Saatavissa: https://www.baal.org.uk/wp-content/uploads/2017/08/goodpractice_full.pdf.

Brothers, Kyle B. & Clayton, Ellen W. 2010. Human non-subjects research: privacy and compliance. *The American Journal of Bioethics* 10:9, 15–17.

Bruckman, Amy. 2002. Ethical guidelines for research online. (Versio: 4/4/02). [verkkoaineisto]. [viitattu 9.4.2019]. Saatavissa: http://documents.kenyon.edu/provost/irb_ethics_online.pdf.

Bruckman, Amy. 2006. Teaching students to study online communities ethically. *Journal of Information Ethics* 15:2, 82–98.

Bruckman, Amy. 2014. Research ethics and HCI. Julkaisussa: Olson, Judith S. & Kellogg, Wendy A. (toim.) *Ways of Knowing in HCI*. New York (NY): Springer, 449–468.

Buchanan, Elizabeth A. 1999. An overview of information ethics issues in a world-wide context. *Ethics and Information Technology* 1:3, 193–201.

Buchanan, Tom. 2000. Potential of the Internet for personality research. Julkaisussa:

Birnbaum, Michael H. (toim.) *Psychological Experiments on the Internet*. San Diego (CA): Academic Press, 121–140.
Cadwalladr, Carole & Graham-Harrison, Emma. 2018. The Cambridge analytica files. *The Guardian* 21, 6–7.
Clarkeburn, Henriikka & Mustajoki, Arto. 2007. *Tutkijan arkipäivän etiikka*. Tampere: Vastapaino.
Costa, Peter I. de (toim.). 2015. *Ethics in Applied Linguistics Research: Language Researcher Narratives*. New York (NY): Routledge.
Creswell, John W. & Creswell, J. David. 2017. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. (5. p.). Thousand Oaks (CA): Sage.
Curtiss, Susan. 1977. *Genie*: *A Psycholinguistic Study of a Modern-Day "Wild Child"*, *Perspectives in Neurolinguistics and Psycholinguistics*. Boston (MA): Academic Press.
Dorian, Nancy C. 1993. A response to Ladefoged's other view of endangered languages. *Language* 69:3, 575–579.
Eaton, Margaret L. & Kennedy, Donald L. 2007. *Innovation in Medical Technology: Ethical Issues and Challenges*. Baltimore (MD): John Hopkins University Press.
Fairfield, Joshua A. 2012. Avatar experimentation: Human subjects research in virtual worlds. *UC Irvine Law Review Vol. 2*, 695. Washington & Lee Legal Studies Paper No. 2012–26.
Fossey, Ellie, Harvey, Carol, McDermott, Fiona & Davidson, Larry. 2002. Understanding and evaluating qualitative research. *Australian and New Zealand Journal of Psychiatry* 36:6, 717–732.
Foucault, Michel. 1982. The subject and power. *Critical Inquiry* 8:4, 777–795.
Gadamer, Hans-Georg. 1976. *Philosophical Hermeneutics*. Kääntänyt David Linge. Berkeley: University of California Press.
Graham, Loren R. 1981. *Between science and values*. New York (NY): Columbia University Press.
Hallamaa, Jaana, Launis, Veikko, Lötjönen, Salla & Sorvali, Irma (toim.). 2006. *Etiikkaa ihmistieteille*. Tietolipas 211. Helsinki: Suomalaisen Kirjallisuuden Seura.
Hudson, James M. & Bruckman, Amy. 2005. Using empirical data to reason about internet research ethics. Julkaisussa: *ECSCW'05 Proceedings of the Ninth Conference on European Conference on Computer Supported Cooperative Work*. Dordrecht: Springer, 287–306.
Hudspith, Robert. 1991. Broadening the scope of engineering ethics: From micro-ethics to macro-ethics. *Bulletin of Science, Technology & Society* 11:4–5, 208–211.
Hunter, David & Evans, Nicholas. 2016. Facebook emotional contagion experiment controversy. *Research Ethics* 12:1, 2–3.
Introna, Lucas D. & Nissenbaum, Helen. 2000. Shaping the Web: Why the politics of search engines matters. *The Information Society* 16:3, 169–185.
Irigaray, Luce. 2017. *To Speak is Never Neutral*. New York (NY): Routledge.
Jones, Peter. 1995. Contradictions and unanswered questions in the Genie case: a fresh look at the linguistic evidence. *Language and Communication* 15, 261–280.
Kimmel, Allan J. 1996. *Ethical Issues in Behavioral Research: Basic and Applied Perspectives*. Oxford: Wiley-Blackwell.

Kitchin, Heather A. 2008. *Research Ethics and the Internet: Negotiating Canada's Tri-Council Policy Statement*. Black Point: Fernwood Publishing Co. Ltd.
Konstenius, Reetta A. 2014. *Empiria, eksperimentti ja etiikka. Kielitieteen metateoriaa*. Helsinki: Helsingin yliopisto.
Kristeva, Julia. 1988. The system and the speaking subject. Julkaisussa: Newton, K. M. (toim.) *Twentieth-Century Literary Theory a Reader*. London: Macmillan, 180–185.
Kubanyiova, Magdalena. 2008. Rethinking research ethics in contemporary applied linguistics: The tension between macroethical and microethical perspectives in situated research. *The Modern Language Journal* 92:4, 503–518.
Lagström, Hanna, Pösö, Tarja, Rutanen, Niina & Vehkalahti, Kaisa. 2015. *Lasten ja nuorten tutkimuksen etiikka*. Helsinki: Yliopistopaino.
Länsimies-Antikainen, Helena. 2008. Ihmisten tutkimiseen liittyviä eettisiä kysymyksiä: pohdintaa tietoon perustuvasta suostumuksesta. Julkaisussa: Pietilä, Anna-Maija & Länsimies-Antikainen, Helena (toim.) *Etiikkaa monitieteisesti. Pohdintaa ja kysymyksiä*. Kuopio: Kuopion yliopisto, 91–108.
Löppönen, Paavo (toim.). 1991. *Tiede ja etiikka*. Porvoo: WSOY.
Lötjönen, Salla. 1999. *Tutkijan ammattietiikka*. Tutkimuseettinen neuvottelukunta. Koulutus- ja tiedepolitiikan osaston julkaisusarja. Nro. 69. Helsinki: Opetusministeriö.
McKee, Heidi A. & Porter, James E. 2009. *The Ethics of Internet Research: A Rhetorical, Case-based Process*. Bern: Peter Lang.
Moor, James H. 1985. What is computer ethics? *Metaphilosophy* 16:4, 266–275.
Munro, Pamela. 2000. Field linguistics. Julkaisussa: Aronoff, Mark & Rees-Miller, Janie (toim.) *The Handbook of Linguistics*. Oxford: Blackwell Publishers, 130–149.
Mäkelä, Klaus. 2005. Sosiaalitutkimuksen eettinen sääntely. *Yhteiskuntapolitiikka* 70:4, 386–395.
Mäkelä, Klaus. 2007. Lääketieteen etiikka ja yhteiskuntatutkimuksen ennakkosäätely. *Tieteessä tapahtuu* 25:1, 21–27.
Niiniluoto, Ilkka & Sihvola, Juha. 2005. *Nykyajan etiikka. Keskusteluja ihmisestä ja yhteisöstä*. Helsinki: Gaudeamus.
Orlitzky, Marc. 2017. Virtue signaling: Oversocialized "integrity" in a politically correct world. Julkaisussa: Orlitzky, Mark & Monga, Manjit (toim.) *Integrity in Business and Management*. London: Routledge, 190–206.
Percival, Thomas. 1803. *Medical Ethics, Or a Code of Institutes and Precepts Adapted to the Professional Conduct of Physicians and Surgeons: To which is Added an Appendix*. Manchester: S. Russell for J. Johnson.
Pietarinen, Juhani & Launis, Veikko. 2002. Etiikan luonne ja alueet. Julkaisussa: Karjalainen, S., Launis, V., Pelkonen, R. & Pietarinen, J. (toim.) *Tutkijan eettiset valinnat*. Helsinki: Gaudeamus, 42–57.
Pirttilä, Tuula. 2008. Tutkimuseettiset toimikunnat – toimintaa ohjaavat normit ja tutkimuseettinen pohdinta. Julkaisussa: Pietilä, A. M. & Länsimies-Antikainen, H. (toim.) *Etiikkaa monitieteisesti. Pohdintaa ja kysymyksiä*. Kuopio: Kuopion yliopisto, 65–89.
Roelcke, Volker & Maio, Giovanni. 2004. *Twentieth Century Ethics of Human Subjects Research: Historical Perspectives on Values, Practices and Regulations*. Stuttgart: Verlag.

Ross, Monique. 2014. Facebook turns 10: The world's largest social network in numbers. *ABC News*. [verkkoaineisto]. [viitattu 14.3.2019]. Saatavissa: http://www.abc.net.au/news/2014-02-04/facebook-turns-10:-the-social-network-in-numbers/5237128.

Rymer, Russ. 1992. *Genie an Abused Child's Flight from Silence*. New York (NY): Harper Collins Publishers.

Tierney, Patrick. 2000. *Darkness in El Dorado: How Scientists and Journalists Devastated the Amazon*. New York (NY): W. W. Norton & Company.

Vaux, Bert & Cooper, Justin. 1999. *Introduction to Linguistic Field Methods*. Munich: Lincom Europa.

Weindling, Paul J. 2008. The Nazi medical experiments. Julkaisussa: Emanuel, Ezekiel, Grady, Christine, Miller, Franklin G., Crouch, Robert A. & Lie, Reidar K. (toim.) *The Oxford Textbook of Clinical Research Ethics*. New York (NY): Oxford University Press, 18–30.

White, Michele. 2002. Representations or people? *Ethics and Information Technology* 4:3, 249–266.

Wiberg, Matti. 2006. Yhteiskuntatiede ja kohteen vahingoittaminen. Julkaisussa: Hallamaa, Jaana, Launis, Veikko, Lötjönen, Salla & Sorvali, Irma (toim.) *Etiikkaa ihmistieteille*. Tietolipas 211. Helsinki: Suomalaisen Kirjallisuuden Seura, 261–273.

Wolfram, Walt. 1993. Ethical considerations in language awareness programs. *Issues in Applied Linguistics* 4:2, 225–255.

Zimmer, Michael. 2010. "But the data is already public": on the ethics of research in Facebook. *Ethics and Information Technology* 12:4, 313–325.

Zimmer, Michael & Proferes, Nicholas John. 2014. A topology of Twitter research: disciplines, methods, and ethics. *Aslib Journal of Information Management* 66:3, 250–261.

# 5. Tämä kokoelma

*Milla Luodonpää-Manni*
https://orcid.org/0000-0002-1646-4462
*Markus Hamunen*
https://orcid.org/0000-0003-2786-6392

## 5.1. Kokoelman rakenne

Tämä teos jakautuu neljään osaan: tähän laajaan, mutta ei tyhjentävään, tutkimuksen perusteisiin ja käytäntöön opastavaan johdantoon (osa 1), kielentutkimuksen aineistonkeruu- ja analyysimenetelmiä käsittelevään osaan 2 (luvut 1–11), kielellisen monimuotoisuuden selvittelyyn soveltuvia menetelmiä esittelevään osaan 3 (luvut 12–16) sekä esimerkkitapauksia menetelmällisestä kirjosta kielentutkimuksen eri alueilla esittelevään osaan 4 (luvut 17–24). Kokoelma sisältää yhteensä 24 kielentutkimuksen eri menetelmiä ja suuntauksia esittelevää lukua, joissa avataan tiivistetysti alan perusteita ja ohjataan lukijaa yksityiskohtaisemman tiedon lähteille. Kirjoittajina toimivat aina kyseisen menetelmän asiantuntijat. Teoksen laatimiseen on osallistunut yhteensä 46 kirjoittajaa. Teos pyrkii kuvaamaan suomalaisen kielentutkimuksen kenttää mahdollisimman laajasti kurkottaen monilta osin kielentutkimuksen tyypillisimpien alueiden (esim. rakenteen ja merkityksen kuvaaminen) ulkopuolelle. Valinta on tietoinen, ja sillä pyritään

lisäämään kielentutkimuksen alueiden välistä dialogia. Tätä dialogia edistetään myös runsailla kokoelman sisäisillä viittauksilla lähialoja esitteleviin kirjoituksiin. Samalla lukijaa perehdytetään tieteenalan rakenteisiin ja tarjotaan lisätietoa käsiteltävästä aiheesta.

Koska kyseessä on ensimmäinen suomenkielinen kielentutkimuksen metodologiaan keskittyvä laaja kokoelma, kirjoituksissa metodeja ja niiden historiaa esitellään erityisesti suomalaisesta näkökulmasta. Metodien kuvaukset on kirjoitettu mahdollisimman konkreettisella ja käytännönläheisellä otteella, ja usein näytetään askel askeleelta esimerkkejä siitä, miten menetelmää käytetään. Kokoelmassa termien tarkkaan määrittelyyn on kiinnitetty erityistä huomiota, koska termien käytössä voi olla eroja kielentutkimuksen eri alueilla. Mahdollisuuksien mukaan käytämme myös suoria ulkoisia linkkejä esimerkiksi Tieteen termipankkiin, mikä on mahdollista teoksen sähköisen julkaisemisen ansiosta. Lukijan työtä pyritään helpottamaan myös visuaalisesti erottuvilla metodien keskeisiä piirteitä esittelevillä tietolaatikoilla ja lukijalle tarjottavalla lisälukemistolla.

Kokoelman johdannossa (osa I) on perehdytty tutkimuksenteon eri vaiheisiin ja pohdittu yleisellä tasolla perustavia tieteen metodologisia ja tutkimusetiikkaan liittyviä kysymyksiä sekä erityisiä kielentutkimuksen menetelmiä koskevia kysymyksiä. Vaikka johdantoon on pyritty keräämään kielentutkimusta yhteisesti koskevia periaatteita, teoksen muut luvut on laadittu siten, että ne toimivat myös itsenäisesti luettuina kuten perinteiset tieteelliset artikkelit. Eri menetelmiin liittyvistä erityiskysymyksistä, kuten tutkimuseettisistä erityispiirteistä, keskustellaan tarvittaessa syvällisemmin kunkin menetelmäluvun yhteydessä. Lisäksi teoksessa on paikoin säilytetty pientä toisteisuutta, jotta luvut toimisivat silläkin tavoin paremmin itsenäisinä kokonaisuuksina.

Kokoelman johdantoa seuraavassa toisessa osassa (luvut 1–11) käsitellään erilaisia aineistotyyppejä, aineistonkeruuseen liittyviä kysymyksiä sekä menetelmiä, joiden avulla aineistoja voidaan analysoida. Osa koostuu luvuista *Kielitaju kielentutkijan työkaluna, Kysely, Haastattelu aineistonkeruun metodina sosiolingvistiikassa, Kokeellinen fonetiikka, Lingvistinen etnografia toimintayhteisöissä, Diskurssianalyysi, Laadullinen aineistopohjainen kielentutkimus, Käsiteanalyysi kielentutkimuksessa,*

*Korpusaineistot*, *Määrällinen korpuslingvistiikka* ja *Historiallinen korpuslingvistiikka*.

Kokoelman kolmannessa osassa (luvut 12–16) esitellään kielellisen monimuotoisuuden selvittelyyn soveltuvia menetelmiä. Kolmannen osan lukuja ovat *Kielikontaktien tutkimus*, *Kielten dokumentointi ja kieliopin kuvaus*, *Kielitypologian menetelmät*, *Historiallinen ja vertaileva kielentutkimus* sekä *Antropologinen lingvistiikka*.

Kokoelman neljännessä osassa (luvut 17–24) tarkastellaan eräitä muita tutkimussuuntauksia menetelmälliseltä kannalta, ja se käsittää luvut *Kieliopin tutkimus ja kielioppiteoriat*, *Näkökulmia filologiseen tutkimukseen*, *Nimistöntutkimus*, *Tekstintutkimus*, *Sosiolingvistiikka*, *Keskustelunanalyysi ja vuorovaikutuslingvistiikka*, *Käännöstiede ja sen menetelmät* sekä *Psyko- ja neurolingvistiikka*.

Kielentutkimuksessa analyysimenetelmät ovat usein sidoksissa tiettyihin tutkimussuuntauksiin, eikä tarkkaa jakoa metodien ja suuntausten kesken käytännössä aina voida tehdä. Tämä käy ilmi myös kokoelman lukujen teemoista (ks. myös kokoelman Alkusanat). Siirtymä yksittäisistä metodeista erilaisiin laajempiin tutkimussuuntauksiin ja monimenetelmäisiin tutkimusasetelmiin on todellisuudessa portaaton, vaikka teoksessa on tehty jako näiden välille teoksen jäsentämiseksi.

Vaikka monet perinteisesti kielentutkimuksessa analysoidut rakenteen ja merkityksen tasot, kuten fonologia (äänneoppi), morfologia (muoto-oppi), semantiikka (merkitysoppi) ja syntaksi (lauseoppi), eivät näy kokoelmassa otsikkotasolla, tämän teoksen luvut tarjoavat kuitenkin monipuolisesti metodisia välineitä myös näiden ja monien muiden kielentutkimuksen kohteiden tarkasteluun. Luvut on järjestetty kokoelmassa löyhien temaattisten periaatteiden mukaan, eikä niiden esittämisjärjestyksellä näin ollen oteta arvottavaa kantaa eri menetelmien ja suuntausten väliseen järjestykseen.

Tämän kokoelman yhtenä tavoitteena on ylittää kielentutkimuksen eri alueiden välisiä raja-aitoja sekä vahvistaa kielentutkimuksen metodologian asemaa osana tieteenalan keskustelukulttuuria, rakenteita ja käytänteitä. Parhaassa tapauksessa teos toimii sysäyksenä metateoreettisten julkaisujen ja toiminnan runsastumiseen suomalaisessa kielentutkimuksessa.

## 5.2. Kielentutkimuksen muita metodioppaita ja perusteoksia suomeksi

Viime aikoina kielentutkimuksen menetelmiä käsitteleviä teoksia on kirjoitettu suomen kielellä jonkin verran. Yleensä ne keskittyvät jonkin tietyn kielentutkimuksen suuntauksen metodiikan kuvaukseen, sillä kielentutkimusta harjoitetaan erilaisten suuntausten piirissä ja niissä erilaisin eriytynein menetelmin.

Tainio (1997) on ensimmäinen suomeksi kirjoitettu **keskustelunanalyysin** perusteita esittelevä oppikirja. Stevanovic ja Lindholm (2016) päivittää keskustelunanalyysin tutkimusprosesseja esittelemällä tuoretta keskustelunanalyyttista tutkimusta, ja teos avaa siten uusia näkökulmia keskustelunanalyysin hyödyntämiseen. Myös jo suomennoksessa Heritage (1996 [1984]) käsitellään keskustelunanalyysin metodologisia perusteita.

Oittinen ja Mäkinen (2001) on ensimmäinen suomeksi kirjoitettu **kääntämisen ja tulkkauksen** perusteita esittelevä yleistajuinen kirja. Oittinen ja Tuominen (2007) tarjoaa puolestaan kattavan yleiskatsauksen av-kääntämisestä. Aaltonen, Siponkoski ja Abdallah (2015) päivittää uusia näkökulmia käännösviestintään keskittyen erityisesti käyttäjäkeskeiseen kääntämiseen.

Ruusuvuori ja Tiittula (2005) on monitieteinen oppikirja erilaisista **haastatteluista**, haastattelutilanteen rakentumisesta sekä menetelmän historiasta. Haastatteluaineiston analyysiin käytännöllisen perehdytyksen tarjoaa puolestaan Ruusuvuori, Nikander ja Hyvärinen (2010). Tutkimushaastattelusta on ilmestynyt myös käsikirja, Hyvärinen, Nikander ja Ruusuvuori (2017), joka opastaa yksityiskohtaisesti haastattelututkimuksen käytäntöön ja esittelee haastattelun eri lajeja. Seuri (2020) avaa haastattelua journalismin näkökulmasta, mistä on hyötyä myös tutkimustyössä.

Mäntynen, Shore ja Solin (2006) on ensimmäinen suomenkielinen **tekstilajeja** käsittelevä teos. Heikkisen, Voutilaisen, Lauerman, Tiililän ja Lounelan (2012a) kattava käsikirja keskittyy niin ikään tekstilajitutkimukseen ja sen käsitteisiin. Sen parina toimii tekstilajien tutkimuksen käytäntöön opastava Heikkinen, Voutilainen, Lauerma, Tiililä ja Lounela (2012b), joka on verkossa avoimesti saatavissa.

Suomi, Toivanen ja Ylitalo (2006) on perusteos suomen kielen **fonetiikasta**, mutta se soveltuu hakuteokseksi monien muidenkin alojen opiskelijoille. Kattava kokoomateos foneettisesta tutkimuksesta on myös Iivonen (2005).

Ainiala, Saarelma ja Sjöblom (2008) on ensimmäinen suomeksi kirjoitettu **nimistöntutkimuksen** perusteita esittelevä oppikirja ja hakuteos.

Itkonen (2008–2010) on kolmiosainen suomenkielinen perusteos **kielitypologiasta**. Siinä esitellään kielitypologian eri osa-alueita.

Karlsson (2015 [2008]) on yleisesitys **kielentutkimuksen perusteista** ja kielen rakenteen tutkimuksen perinteisistä osa-alueista (esim. fonetiikka, fonologia, morfologia). Siinä on myös tiivis luku kielitieteen metodiikasta. Myös Häkkinen (2007) on yleisen kielitieteen perusteos, joka sopii kaikille kielitieteen teoriasta kiinnostuneille.

Idström ja Sosa (2009) on monipuolinen kokoelma **kielten ja kulttuurien vuorovaikutukseen** keskittyviä artikkeleita.

Ojutkangas, Larjavaara, Miestamo ja Ylikoski (2009) tarjoaa merkityslähtöisen esityksen **kielestä ja kielentutkimuksesta**. Teos käsittelee kieliä kielitypologisesta, maailmanlaajuisesti kieliä vertailevasta näkökulmasta.

Pietikäinen ja Mäntynen (2009) on **diskurssintutkimuksen** perusteos, joka avaa havainnollisesti diskurssintutkimuksen käsitteitä ja sovelluksia. Ajantasaistettu laitos tästä on Pietikäinen ja Mäntynen (2019). Jokisen, Juhilan ja Suonisen (2016) teos yhdistää aiemmat oppikirjat *Diskurssianalyysin aakkoset* ja *Diskurssianalyysi liikkeessä*.

Hiekkalinna ja Seppänen (2010) esittelee **klassisen filologian** lähteitä, lähestymistapoja ja metodeja.

Itkonen ja Pajunen (2010) tarkastelee **empiirisen kielitieteen** kenttää ja kielentutkimuksessa käytettyjä metodeja. Se argumentoi metodien välisten edellyttämyssuhteiden puolesta.

Korpilahti, Aaltonen ja Laine (2010) on monitieteinen oppikirja, joka antaa vankan tiedollisen pohjan **kielen ja aivojen yhteistoiminnasta**. Teos esittelee muun muassa puheen ja kielen kehityksen eri vaiheita, kielen häiriöitä ja kuntoutuksellisia ratkaisuja. Puheterapiaan vuorovaikutuksellisesta näkökulmasta pureutuu Sellman ja Tykkyläinen (2017).

Kalaja, Alanen ja Dufva (2011) on **tutkielman laadintaan tarkoitettu opas**, joka soveltuu soveltavan kielentutkimuksen ja erityisesti kielen oppimisen tutkimuksen opiskelijoille.

Kolehmaisen, Miestamon ja Nordlundin (2013) kattava ja uusia innovaatioita sisältävä teos keskittyy erityisesti **vertailun menetelmiin** kielentutkimuksen eri alueilla.

Vehkalahti (2019) on käytännöllinen opas erityisesti **kyselyjen** suunnitteluun ja toteuttamiseen, mutta siinä avataan myös tilastollisia menetelmiä ja käsitteitä.

## MUUTA PERUSKIRJALLISUUTTA:

Lakomäki, Sami, Latvala, Pauliina & Lauren, Kirsi. 2011. *Tekstien rajoilla – monitieteisiä näkökulmia kirjoitettuihin aineistoihin*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Lappalainen, Sirpa, Hynninen, Pirkko, Kankkunen, Tarja, Lahelma, Elina & Tolonen, Tarja (toim.). 2008. *Etnografia metodologiana. Lähtökohtana koulutuksen tutkimus*. Tampere: Vastapaino.

Piippo, Irina, Vaattovaara, Johanna & Voutilainen, Eero. 2016. *Kielen taju: vuorovaikutus, asenteet ja ideologiat*. Helsinki: Art House.

Sajavaara, Kari & Piirainen-Marsh, Arja (toim.). 2009. *Kieli, diskurssi ja yhteisö. Soveltavan kielentutkimuksen teoriaa ja käytäntöä*. Jyväskylän yliopisto: Soveltavan kielentutkimuksen keskus.

## KIRJALLISUUS

Aaltonen, Sirkku, Siponkoski, Nestori & Abdallah, Kristiina (toim.). 2015. *Käännetyt maailmat. Johdatus käännösviestintään*. Helsinki: Gaudeamus.

Ainiala, Terhi, Saarelma, Minna & Sjöblom, Paula. 2008. *Nimistöntutkimuksen perusteet*. Tietolipas 221. Helsinki: Suomalaisen Kirjallisuuden Seura.

Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä, Ulla & Lounela, Mikko (toim.). 2012a. *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Kotimaisten kielten keskuksen julkaisuja 169. Helsinki: Gaudeamus.

Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä, Ulla & Lounela, Mikko (toim.). 2012b. *Genreanalyysi – tekstilajitutkimuksen käytäntöä*. [verkkoaineisto]. [viitattu 18.4.2019]. Kotimaisten kielten keskuksen verkkojulkaisuja 29. Helsinki: Kotimaisten kielten keskus. Saatavissa: http://scripta.kotus.fi/www/verkkojulkaisut/julk29/.

Heritage, John. 1996 [1984]. *Harold Garfinkel ja etnometodologia*. Suomentaneet Ilkka Arminen, Outi Paloposki, Anssi Peräkylä, Sanna Vehviläinen & Soile Veijola. Helsinki: Gaudeamus.

Hiekkalinna, Tiina & Seppänen, Minna. 2010. *Philologia classica*. Lähteitä, lähestymistapoja ja metodeja. Turku: Turun yliopisto.

Hyvärinen, Matti, Nikander, Pirjo & Ruusuvuori, Johanna (toim.). 2017. *Tutkimushaastattelun käsikirja*. Tampere: Vastapaino.

Häkkinen, Kaisa. 2007. *Kielitieteen perusteet*. Tietolipas 133. Helsinki: Suomalaisen Kirjallisuuden Seura.

Idström, Anna & Sosa, Sachiko (toim.). 2009. *Kielissä kulttuurien ääni*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Iivonen, Antti (toim.). 2005. *Puheen salaisuudet. Fonetiikan uusia suuntia*. Helsinki: Gaudeamus.

Itkonen, Esa. 2008–2010. *Maailman kielten erilaisuus ja samuus*. Osat 1–3. Yleisen kielitieteen julkaisuja 12–14. Turku: Turun yliopisto.

Itkonen, Esa & Pajunen, Anneli. 2010. *Empiirisen kielitieteen metodologia*. Suomi 199. Helsinki: Suomalaisen Kirjallisuuden Seura.

Jokinen, Arja, Juhila, Kirsi & Suoninen, Eero. 2016. *Diskurssianalyysi*. Tampere: Vastapaino.

Kalaja, Paula, Alanen, Riikka & Dufva, Hannele. 2011. *Kieltä tutkimassa. Tutkielman laatijan opas*. Helsinki: Finn Lectura.

Karlsson, Fred. 2015 [2008]. *Yleinen kielitiede*. (5. p.). Helsinki: Gaudeamus.

Kolehmainen, Leena, Miestamo, Matti & Nordlund, Taru. 2013. *Kielten vertailun metodiikka*. Suomalaisen Kirjallisuuden Seuran Toimituksia 1387. Helsinki: Suomalaisen Kirjallisuuden Seura.

Korpilahti, Pirjo, Aaltonen, Olli & Laine, Matti (toim.). 2010. *Kieli ja aivot*. Turku: Turun yliopisto.

Mäntynen, Anne, Shore, Susanna & Solin, Anna (toim.). 2006. *Genre – tekstilaji*. Tietolipas 213. Helsinki: Suomalaisen Kirjallisuuden Seura.

Oittinen, Riitta & Mäkinen, Pirjo (toim.). 2001. *Alussa oli käännös*. Tampere: Tampere University Press.

Oittinen, Riitta & Tuominen, Tiina (toim.). 2007. *Olennaisen äärellä. Johdatus audiovisuaaliseen kääntämiseen*. Tampere: Tampereen yliopisto.

Ojutkangas, Krista, Larjavaara, Meri, Miestamo, Matti & Ylikoski, Jussi. 2009. *Johdatus kielitieteeseen*. Helsinki: WSOY.

Pietikäinen, Sari & Mäntynen, Anne. 2009. *Kurssi kohti diskurssia*. Tampere: Vastapaino.

Pietikäinen, Sari & Mäntynen, Anne. 2019. *Uusi kurssi kohti diskurssia*. Tampere: Vastapaino.

Ruusuvuori, Johanna & Tiittula, Liisa (toim.). 2005. *Haastattelu. Tutkimus, tilanteet ja vuorovaikutus*. Tampere: Vastapaino.

Ruusuvuori, Johanna, Nikander, Pirjo & Hyvärinen, Matti (toim.). 2010. *Haastattelun analyysi*. Tampere: Vastapaino.

Sellman, Jaana & Tykkyläinen, Tuula. 2017. *Puheterapia. Vuorovaikutus muutoksen välineenä*. Tampere: Vastapaino.

Seuri, Olli. 2020. *Avoin kysymys. Haastattelun käsikirja*. Tampere: Vastapaino.

Stevanovic, Melisa & Lindholm, Camilla (toim.). 2016. *Keskustelunanalyysi. Kuinka tutkia sosiaalista toimintaa ja vuorovaikutusta*. Tampere: Vastapaino.
Suomi, Kari, Toivanen, Juhani & Ylitalo, Riikka. 2006. *Fonetiikan ja suomen äänneopin perusteet*. Helsinki: Gaudeamus.
Tainio, Liisa. 1997. *Keskustelunanalyysin perusteet*. Vastapaino: Tampere.
Vehkalahti, Kimmo. 2019. Kyselytutkimuksen mittarit ja menetelmät. Helsinki: Helsingin yliopisto. Saatavissa: http://doi.org/10.31885/9789515149817.

# Osa II
# Aineistonkeruusta analyysiin

# Luku 1 Kielitaju kielentutkijan työkaluna

*Markus Hamunen*
https://orcid.org/0000-0003-2786-6392

*Tuomas Huumo*
https://orcid.org/0000-0002-6589-0732

**Mikä?**

*Kielitaju* on kokoava nimitys erilaisille metakielellisille taidoille ja tiedoille. Sillä tarkoitetaan ennen kaikkea äidinkielisen puhujan ymmärrystä omasta kielestään. Se koskee muun muassa kielenmukaisten ja kielenvastaisten ilmausten tunnistamista. Jonkinlainen kielitaju on kaikilla kielenpuhujilla omasta äidinkielestään, mutta myös vierasta kieltä opittaessa syntyy sitä koskeva kielitaju. Se kehittyy vuorovaikutuskokemusten, reflektoinnin ja osin myös formaalin koulutuksen myötä. Kielentutkimuksen metodina kielitaju on erikseen harjaannutettava taito, jota voi käyttää joko sellaisenaan tai muiden metodien, esimerkiksi korpustutkimuksen, tukemana. Kielitaju toimii välttämättömänä edellytyksenä monen muun metodin taustalla. Tavallisin kielitajun sovellusalue on ollut erilaisten kielioppien kirjoitus sekä teoreettinen kielentutkimus.

**Katso myös:**

Johdannon luku 3. Monimenetelmäisen kielentutkimuksen metateoreettisia lähtökohtia

Luku 2 Kysely

Luku 9 Korpusaineistot

Luku 13 Kielten dokumentointi ja kieliopin kuvaus

Luku 17 Kieliopin tutkimus ja kielioppiteoriat

# 1. Johdanto

Suomea äidinkielenään (ensikielenään) puhuva tietää, että ilmaus *tämä punainen talo* on kielessä hyväksyttävä, kun taas **talo punainen tämä* ei ole.[1] Mihin tieto perustuu? Mistä juontuu tuntemamme varmuus siitä, että ilmauksista vain ensimmäinen on hyväksyttävä ja toinen kielenvastainen? Entä miksei jälkimmäinen ilmaus muutu kertaheitolla hyväksyttäväksi, vaikka kirjoittaisit sen tuhat kertaa erilaisille internetin keskustelupalstoille ja väittäisit sitten, että kyllä se suomea on, koska se esiintyy aidossa kielenkäytössä?[2]

Äidinkieltä koskeva tieto, jota usein kutsutaan *kielitajuksi*, kohdistuu puhujayhteisölle yhteiseen normijärjestelmään, jota yksi puhuja ei pysty mielivaltaisesti muuttamaan. Kieltä opetteleva lapsi omaksuu tiedon tästä järjestelmästä ja sitä myötä kielitajun, joka edellä kuvatun kaltaisissa selvissä tapauksissa kertoo, mikä kielessä on mahdollista ja mikä ei. Kielen sääntöjä on usein verrattu pelin, esimerkiksi shakin, sääntöihin: tietynlaiset siirrot ovat pelin sääntöjen mukaisia, toiset eivät. Säännöt ovat pelaajien yhteisesti sisäistämät ja eri pelaajille samat. Niitä voi rikkoa, mutta tällöin pelaaja tekee virheen, ja vastustaja todennäköisesti

protestoi. Normijärjestelmän yhteisyys mahdollistaa sen, että kielellinen kommunikointi ylipäänsä voi onnistua. (Normatiivisuudesta ks. Mäkilähde, Leppänen & Itkonen 2019.)

Kielitaju on myös kielentutkijan välttämätön ja väistämätön työkalu, jota tutkija voi käyttää apunaan karkeasti ottaen kolmessa tutkimuksen eri vaiheessa: 1) tutkimusongelman hahmottamisessa ja tutkimuksen suunnittelussa, 2) tutkimusaineiston tuottamisessa esimerkkejä sepittämällä ja 3) tutkimusaineiston analysoinnissa. Suurin osa kielentutkimuksesta, myös tiukasti korpusten käyttöön pohjautuva empiirinen tutkimus, hyödyntää kielitajua ainakin kohdassa 1 ja usein myös kohdassa 3, kun empiirisistä havainnoista tehdään kvalitatiivisia päätelmiä. Sitä vastoin kohta 2 erottaa kielentutkimuksen suuntauksia toisistaan, ja sen luotettavuudesta on viime vuosikymmeninä käyty melko paljon keskustelua (fennistiikassa mm. Huumo 2007; Ojutkangas 2008; Jantunen 2009). Kohta 2 edellyttää yleensä sitä, että tutkija on tutkittavan kielen äidinkielen tasoinen puhuja. Tutkimuksen aineistona on yleensä standardoitu kielimuoto, esimerkiksi suomen yleiskieli, mutta huomattavasti harvemmin vapaa puhekieli tai esimerkiksi aluemurre, joissa esiintyy paljon variaatiota ja joita koskevat intuitiot eivät siksi ole yhtä vahvoja kuin jaettua yleiskieltä koskevat.[3]

Kielitajun käyttöä tutkimuksessa ja etenkin siihen nojautumista myös tutkimusaineiston tuottamisessa (kohta 2 edellä) on kutsuttu toisinaan *intuitioksi* ja toisinaan *introspektioksi* tarkoittaen samaa asiaa (Willems 2012, 671–672; ks. myös Devitt 2014; Maynes & Gross 2013). Käytämme tässä artikkelissa pääsääntöisesti termiä *kielitaju* tekemättä tarkkaa terminologista eroa lähikäsitteiden kesken. Kielitaju kohdistuu puhujayhteisölle yhteiseen normiin, joka ei ole yksilöstä riippuvainen. Se seikka, että **talo punainen tämä* ei suomessa ole hyväksyttävä, on puhujayhteisön yhteistä tietoa eikä yksittäisen puhujan mielipide tai mielle. Jokaisella puhujalla on kyllä yksilötasolla tieto tästä normista, mutta siihen liittyy samalla tieto siitä, että normi ei ole yksilökohtainen vaan koskee koko yhteisöä (yksilön kieltä koskevasta tiedosta ks. esim. Mills [ei vuotta]; Ludlow 2009). Voin esimerkiksi olettaa, että äidinkielinen keskustelukumppanini oudoksuisi ilmausta *talo punainen tämä*, mikäli sellaista käyttäisin.

Leonard Talmy (2007a) on verrannut kielitajuaan käyttävää kielitieteilijää urheilijaan, joka harjoittaa määrätietoisesti fyysisiä kykyjään. Samoin kielentutkija harjoittaa ja harjaannuttaa omaa kielitajuaan havaitsemaan erilaisia säännönmukaisuuksia ja hienoja merkitysvivahteita, joita tavallinen puhuja ei välttämättä tiedosta tai osaa kielentää mutta joista hänkin voi tulla tietoiseksi, jos ne osoitetaan hänelle. Tutkittavaa kieltä äidinkielen tasoisesti hallitseva tutkija on tällöin luonnollisesti vahvoilla verrattuna tutkijaan, jolle tutkimuskohde on vieras kieli. Ero tulee erityisen vahvana esille, jos myös tutkittava aineisto on tarkoitus tuottaa tutkijan oman kielitajun avulla. Erinomaisestikin vierasta kieltä hallitsevan tutkijan on syytä suhtautua varovasti ilmausten itse muotoilemiseen, ja tällaisten ilmausten hyväksyttävyyttä ja merkitysvivahteita on aina syytä kysyä myös äidinkielisiltä puhujilta (jolloin tutkimuksen tukena käytetään kielenopasta; ks. Kielten dokumentointi ja kieliopin kuvaus tässä kirjassa). Myös äidinkielisen tutkijan on syytä olla varovainen silloin, kun itse muotoillut ilmaukset ovat esimerkiksi rakenteellisesti hyvin kompleksisia tai muuten erikoislaatuisia. Mitä kauemmas etäännytään ns. selvistä tapauksista, sitä epävarmemmin kielitaju kertoo, mikä ilmauksen merkitys on tai missä kulkee hyväksyttävän ja ei-hyväksyttävän raja. Esimerkkejä tällaisen tutkimuksen ongelmallisuudesta tarkastelee havainnollisesti Dąbrowska (1996).

Mahdollisen ja mahdottoman tai hyväksyttävän ja ei-hyväksyttävän ilmauksen välisten erojen pohtimisen lisäksi kielitaju on korvaamaton työkalu silloin, kun pohditaan ilmausten merkityseroja: onko esimerkiksi jonkin kielenaineksen esiintymä tietyssä kontekstissa merkitykseltään erilainen kuin toisenlaisessa kontekstissa? Johtaako ilmauksen yhden elementin korvaaminen toisella merkityksen muuttumiseen, ja jos johtaa, niin millaiseen? Tällainen tutkimus operoi **minimipareilla**, joissa muunnellaan kielellisiä ilmauksia yleensä vain yhden piirteen suhteen ja tarkastellaan muunnoksen vaikutusta ilmauksen merkitykseen (esim. *kettu kuoli kaivo**ssa** ~ kaivo**on***). Minimiparit ovat kontekstittomia, tutkijan itsensä konstruoimia ilmauksia, joiden tavoitteena on eristää parikkeja erottavan tekijän vaikutus niiden merkitykseen. Samoin kuin hyväksyttävyyttä koskeva intuitio myös tämä menetelmä toimii parhaiten silloin, kun havaittu merkitysero on selkeä ja yksiselitteinen.

Yksinomainen kielitajun ja erityisesti tutkijan itse konstruoimien esimerkkien varassa operoiva kielentutkimus on viime vuosikymmenten aikana joutunut kritiikin kohteeksi (esim. Schütze 1996). Osin kritiikki on ollut perusteltua, osin perustelematonta siksi, että kielitajun varaisen tutkimuksen on oletettu pyrkivän sellaisiinkin tavoitteisiin, joihin se todellisuudessa ei pyri. Myös kritiikki, jossa esitetään kokonaan kielitajun käytöstä luopumista kielentutkimuksessa, kohtaa todellisia haasteita, sillä kielentutkija ei käytännössä voi välttää kielitajunsa käyttöä.

Esittelemme tässä artikkelissa tutkijan kielitajun käyttöä kielitieteellisenä metodina sekä tieteenfilosofiselta että käytännölliseltä kannalta. Kohteeksi rajautuu pääasiassa tutkijan äidinkieli. Luvussa 2 esitämme ensin kielitajun tietoteoreettisen perustan ja sitten vertailemme ei-kielentutkijan ja kielentutkijan kielitajua. Luvussa 3 esittelemme kielitajulle läheisiä käsitteitä (ks. myös tietolaatikko 1). Luku 4 esittelee tiiviisti kielitajun käytön historiaa kielentutkimuksessa ja kokoaa siihen kohdistettua kritiikkiä. Luvussa 5 havainnollistamme kielitajun käyttöä käytännössä erikseen huomioitavien seikkojen ja esimerkkitapauksen valossa. Luku 6 on yhteenveto.

# 2. Mitä on kielitaju?

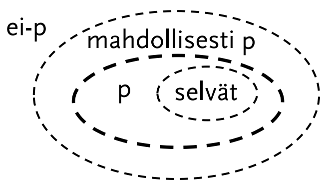
## 2.1. Mikä on kielitajun tiedollinen perusta?

Kielitaju kohdistuu siis kieltä koskeviin tietoihin ja taitoihin. Kysyttäessä, mitä tällaiset tiedot ja taidot ovat tai voivat olla, nousee luonnostaan tarkasteltavaksi myös kysymys, miten tällaiset asiat periaatteessa tiedetään tai miten niihin on kielenkäyttäjällä pääsy (ks. luku 2.2.). Edellytys kielitajun täysipainoiselle käytölle on äidinkielisyys tai muulla tavoin hankittu ensikielen tasoinen kielitaito (Karlsson 1998, 287). Usein tarkoitetaan äidinkielensä rakenteelliset perusteet jo oppinutta puhujaa.[4] Tämän edellytyksen täyttyessä kielenpuhuja P voi todeta, että esimerkki (1b) on suomea ja esimerkit (1a) ja (1c) eivät ole.[5] P on saattanut kuulla

jossain lausuman (1a). Hän kykenee sen palauttamaan mieleensä, on kielitajunsa varassa varma siitä, ettei kyse ole suomesta, mutta ei viroa taitamattomana tiedä, mitä se tarkoittaa. Viroa osaavana P voi olla myös toisella tapaa vakuuttunut, että kyseessä ei ole suomi, ts. koska hän tietää (1a):n olevan viroa. Esimerkki (1a) voi muistuttaa P:n mielessä esimerkkiä (1b). Tässä suhteessa vieläkin riidattomampi suomenmukaisuus lienee esimerkki (1d), joka edustaa ns. selvää tapausta (engl. *clear case*, ks. Karlsson 1998, 287; Itkonen 2003, 34). Vastaavasti (1c) ei selvästi ole suomea.

(1a) *Kas sa oled täna heas tujus?* ’Oletko tänään hyvällä tuulella?’ [viroa]
(1b) *Kas, sä olet tänään hyväs tujus.* ’Kas vain, oletpa tänään tukevassa humalassa.’ [puhekielinen ilmaus]
(1c) **Koskantti, sokäntti koletontti tonoonkäntti hovaskyntti tojoskuntti* [kontinkieli, leikkikieli]
(1d) *Naapuri on tukevassa humalassa.*

Esimerkeillä (1a–d) voidaan rajata summittaisesti niitä tiedonaloja, joihin kielitajua voidaan pätevästi soveltaa. Kuviossa 1 kielitajun kannalta selkeimmät alueet ovat p ja ei-p (p ’käypä ilmaus’): Jälkimmäinen rajaa selvästi P:n kielitajun ulottumattomiin hänelle vieraat luonnolliset kielet ja keinotekoiset hölynpöly- tai privaattikielet (vrt. sosiaalisesti rajatut leikki- ja koodikielet), joita P ei hallitse. Ala p on yleensä kielitajulle kielenkäytön kannalta selvemmin tarjolla kuin ala ei-p, joka ilmenee




Kuvio 1. Äidinkieltä koskevan kielellisen tiedon loogisia aloja. (P:n mielestä p, missä p = ’käypä ilmaus eli jokin ilmaus on esim. suomea tai jokin ilmiö koskee esim. suomen kieltä’. Huom. Alueiden koot ja suhteet viitteellisiä.)

tyypillisesti metakielellisissä pohdinnoissa ("onko oikein tai kielenmukaista sanoa näin?"). Luvussa 2.2. käsittelemme tarkemmin käypiä ilmauksia. Luvussa 4.1. käsittelemme esimerkkejä kielitajun raja-alueesta sisällyttäen mukaan kielitajun kannalta vaikeasti tai ei ollenkaan tavoitettavia ilmauksia.

Kun tarkastellaan sitä, millaisia kielellisiä ilmiöitä P:n kielitajulle luultavimmin selvinä tapauksina näyttäytyvät asiat ovat, voidaan esittää muun muassa esimerkkien (2) ja (3) tapaisia seikkoja. Ne koskevat kaikki jollain tavoin kielen käyttöä. Esimerkit koskevat eräitä suomen normeja.

(2a) "/p/ lausutaan tietyn artikulatorisen variaation sallimissa rajoissa [p] sanassa *palo*" [/p/ foneemi ja [p] sen toteutuma äänteenä]

(2b) *ty**rs**ky* vs. **ty**sr**ky*

(2c) *talo+on* vs. **on+talo* [*talo*-INE] tai *korpi* : *korve+n* vs. **korpi+n* [korpi-GEN]

(2d) *sen naapurin kivat pikku kekkerit* vs. **naapurin pikku se kivoja kekkerit*

(3a) [$P_1$: *Huomenta*!] $P_2$: *Huomenta vaan*! vs. $P_2$: **Alasti tietysti*!

(3b) *Minulla oli joskus **pikku mopo**. Ja **se** oli mielenkiintoinen vempain.*

(3c) *Naapuri on taas "ihan vain pienissä".* [lainausmerkeissä oleva jakso ironisesti]

Esimerkki (2a) koskee fonologiaa ja foneemin toteutumista äänteenä. Foneettisen variaation rajoissa /p/:n toteutumat kyseisessä foneettisessa kontekstissa antavat äännejakson [palo], joka on tulkittavissa esimerkiksi sanaksi *palo* 'palo' (esim. *maasto**palo***, ***palo**kunta*, vrt. ***Palo** Alto*). Tulkinta koskee semantiikkaa. Esimerkki (2b) koskee fonotaksia eli äänteiden lineaarista järjestystä: jono *rs* on suomessa mahdollinen, **sr* ei ole missään kontekstissa yhdistämättömissä sanoissa (vrt. *hiu**sr**aja*). Esimerkki (2c) havainnollistaa morfologiaa: ensimmäinen morfotaksia eli morfeemien järjestystä ja jälkimmäinen morfofonologiaa eli morfeemien yhdistämisestä aiheutuvia fonologisia muutoksia (tässä suomen

astevaihtelua sanaa taivutettaessa). Esimerkki (2d) koskee syntaksia eli sanojen keskinäistä yhdistämistä (sanojen järjestystä ja kongruenssia).

Esimerkit (2a–d) koskevat pääsääntöisesti kielen rakennetta tai muotoa (ks. myös Karlsson 1982a, 18–28). Kielitajulla tavoitetaan kuitenkin myös erilaisia merkitykseen ja kielen käyttöön kuuluvia seikkoja. Esimerkki (3a) koskee vuorovaikuttamista: vuoroparin jälkijäsen voi, vuorovaikutuksen mielekkyyden säilyessä, vaihdella vain tietyssä määrin. Vuoropareja ovat keskustelun rutiineista esimerkiksi tervehdys–tervehdys ja kysymys–vastaus-parit (ks. Keskustelunanalyysi ja vuorovaikutuslingvistiikka tk.). Esimerkki (3b) liittyy käynnissä olevan tekstin topiikin tunnistettavaksi tekemiseen ja siihen viittaamiseen sopivalla pronominilla. Esimerkki (3c) havainnollistaa kykyä tulkita esimerkiksi toisen puheesta viestinnällisiä epäsananmukaisuuksia tai muodostaa sellaisia itse.

Kuten esimerkeistä (2) ja (3) havaitaan, kielitajun kohteita voi olla periaatteessa kaikilta kielenkuvauksen perinteisiltä kuvaustasoilta aina äänne- ja muoto-opista pragmatiikkaan asti. On tärkeää korostaa muutamia asioita. Ensiksi vaikka edellä esitetyt ilmiöt kuvaavat erilaisia kieltä koskevia yksittäisseikkoja, niistä ei kielenpuhujalle muodostu irrallista faktojen listaa. Sen sijaan monet yksittäiset tiedot linkittyvät toisiinsa muodostaen kieliopin järjestelmän tai osajärjestelmiä. Ei ole ylipäätään selvää, kuinka paljon erilaisia normatiivisia yksittäisseikkoja kielenpuhuja hallitsee tai tiedostaa, mutta perinteiset kielioppien kokonaiskuvaukset ovat tässä suhteessa ainakin suuntaa antavia. Esimerkiksi (2a):ssa *p*-äänteen erilaiset artikulatoriset mahdollisuudet, eli miten *p*:n voi ääntää, rajautuvat muihin äänteisiin (esim. ***v**alo*, ***t**alo*), ja tämän varaan rakentuu viimein tieto suomen koko äännejärjestelmästä. Samoin esimerkki (2c) ei havainnollista vain *talo*-sanan yksikön illatiivin muodostusta tai vain *korpi*-sanan astevaihtelua, vaan erilaiset taivutuksen periaatteet koskevat samalla tavoin kaikkia vastaavanlaisia nomineja kaikissa sijoissa.

Toiseksi tiettyjen ei-kielenmukaisuuksien tuottaminen on kielen oppimisen jälkeen sekä eriasteista että epätodennäköisempää kuin toisten. Esimerkiksi *korpi*-sanan yksikön genetiivi on kielitajussa mitä todennäköisimmin *korven*, ei **korpin* (2c). Vastaavasti taas jokin tervehdykseen vastaaminen (3a) voi tapahtua huomattavasti vapaammissa rajoissa (esim. 3a; $P_2$: *Katohan, Keijo* ~ kädenheilautus tai muu ele ~

ei reagointia). Tämä johtuu siitä, että esimerkiksi monet muoto-opin ilmiöt ovat rajatumpia kuin pragmatiikan ilmiöt. Esimerkiksi jokin taivutustyyppi voi koskea jotain lueteltavaa sanajoukkoa. Normeilla on eri luonne kielen eri tasoilla.

Kolmantena ja tärkeimpänä huomiona voidaan todeta, että arkisessa kielenkäytössä kielitaju toimii sekä yksittäisissä kielenkäyttötilanteissa (*online-introspektio*) että käytöstä irrallisena kognitiivisena kykynä (*offline-introspektio*) (Talmy 2007a; 2007b; ks. luku 5.2.). Perinteisesti kielitajua on tutkimuksessa sovellettu ja se on myös ymmärretty jälkimmäisellä tavalla.

## 2.2. Kielitajun varaisen tietämisen perusteita ei-kielentutkijalla ja kielentutkijalla

Kielitaju on kielen oppimisen ja kulttuuriin sosiaalistumisen seuraus. Ihmisillä oman ja toisten kielen erilaista monitorointia tapahtuu kielenkäyttötilanteissa ja niistä irrallaan kaiken aikaa. Kielenpuhujan ja kielentutkijan välillä on tässä suhteessa kuitenkin havaittavia motivaatio-, tiedostumis-, tavoite- ja muita laadullisia eroja.[6] Ensinnä kielenpuhujan motivaatiosuhde kieleen on samankaltainen kuin sekä kaivajan suhde lapioon että taiteilijan suhde pensseliin: kieltä käytetään; se on käyttöväline maailmassa ja sen avulla toimitaan erilaisissa vuorovaikutustilanteissa. Samoin kielen monitoroinnin tavoitteet koskevat pääosin sitä, miten erilaisissa tilanteissa selvitään kielellisesti, miten keskinäinen (epä)ymmärrys ratkaistaan, miten ihmissuhteita ylläpidetään jne. Toiseksi omaan ja toisten kieleen sekä kanssakäymiseen kohdistuu erilaisia tunteita ja suhtautumistapoja, mikä ilmenee esimerkiksi kielivitseissä, murrematkimuksissa ja oikeakielisyyskeskusteluissa. Käytöstä irrallista itsehavainnointiin perustuvaa monitorointia tapahtuu kaikilla, ja esimerkiksi koulujärjestelmän kieltä yleisesti ja (vieraita) kieliä erityisesti koskeva opetus myös kehittää tällaisen kyvyn harjoittamista.[7] Kouluopetuksen kieleen kohdistama huomio on jo tiedeperustaista.

On ehkä mahdotonta todentaa täysin varmasti, millainen on kielenpuhujan keskimääräinen kielitaju (ks. Dąbrowska 2010), mutta

Wittgensteinin teos *Varmuudesta* (1999 [1969]) tarjoaa uskottavia filosofisia lähtökohtia kielitajun ymmärtämisen perustaksi:

> Kummallista tässä on se, että vaikka olen aivan varma sanojen [kielen] käytöstä, vaikka minulla ei ole siitä mitään epäilyä, en silti voi ilmoittaa menettelytavalleni mitään *perusteita*. Jos yrittäisin, voisin esittää 1000 perustetta, mutta en yhtään perustetta, joka olisi yhtä varma kuin se, mitä perusteiden on juuri määrä perustella. (§ 307)
>
> [– –]. He [ihmiset] ovat oppineet aina kokemuksesta ja heidän toimistaan voidaan nähdä, että he uskoivat varmasti tiettyjä asioita [esim. kieltä koskien], lausuivat he tämän uskon tai eivät. Tällä en tietenkään halua sanoa, että ihmisten *pitää* toimia näin, ainoastaan, että he toimivat näin. (§ 284)

Triviaalistaen ja kieleen soveltaen Wittgensteinin ajatusta kielenpuhujan itsetiedostavasta kielenkäytöstä voi havainnollistaa seuraavilla esimerkeillä (Wittgenstein 1999 [1969], § 28):

(4) *Tämä*$_{\textit{talo}\text{-sanan yksikön illatiivi}}$ tapahtuu$_{\text{taivutetaan}}$ *näin*$_{\textit{taloon}}$. [vrt. 2c]

(5) *Tapahtuu niin*, että tämä$_{\text{/r/}}$ tulee tätä$_{\text{/s/}}$ ennen tässä$_{\text{sanassa } \textit{tyrsky}}$. [vrt. 2b]

Kieli on käyttöä (mt. § 110, 204, 229), ja tapahtumiskuvaukset (4) ja (5) koskevat varmuuden kyllästäminä kielenpuhujan mielessä kaikkia yksittäisiä selviä tapauksia.[8] Näitä koskeva varmuus tai vakuuttuneisuus on pääosin piilevää – niiden perusteita ei yleensä esitetä –, ikään kuin kaiken kokonaisvaltaisen toiminnan perustana ja taustalla oleva episteeminen asenne, jota voi kutsua myös *intuitioksi* (ks. Raami 2015, 21–23; Chudnoff 2013; Koksvik 2011). Tavallisen kielenpuhujan (ei-kielentutkijan) vakuuttuneisuus on suoraan nähtävissä itse kielellisestä toiminnasta ja käytännöstä (Wittgenstein 1999 [1969], § 174, 395). Toiminnan taustalla oleva perusta tai esitieto lienee valtaosin intuitiivisena kielentämätöntä ja ei-propositionaalista eli ei-lausemuotoista. Se vain ilmenee – Wittgensteinin sanoin – *niin* ja *niin* toimimisena. Tätä

voidaan intro- ja retrospektiivisesti saattaa propositionaaliseen muotoon tarkasteltavaksi (esim. 4 ja 5).

Edellä esitettyjä kieltä koskevia varmoja tapauksia edustavat kuvaukset (4) ja (5) ovat naiivia kielitajun eksplikointia ja itse käytännöstä irrotettua ja etäännytettyä toimintaan kohdistuvaa objektivoivaa ajattelua. Itkonen (1974, 76) on kutsunut tämänkaltaisia propositionaalisia kuvauksia *esiteoreettisiksi* (*ateoreettisiksi*) *sääntölauseiksi*. Kielitajun varmuus arkisen kielellisen toiminnan taustalla ilmenee kuitenkin monina muina epäsuorina seikkoina: Emme epäile varmoja tapauksia (esim. 4: ”Tapahtuuko *tämä* tosiaan *näin*? Onko *näin* aina tapahtunut?”), tuskin erehdymme niissä ja erehtymisen mahdollisuus ei tavallisesti edes nouse mieleemme (Wittgenstein 1999 [1969], § 38, 83, 155). Uudet ulkoiset havainnot tai löydökset varmoina pitämistämme asioista eivät tee meitä entistä vakuuttuneemmiksi niistä tai siitä, mitä erehtyminen voisi ylipäätään näissä tapauksissa olla (ks. mt. § 307). Lisäksi emme pidä varmuuden esiteoreettista muotoilua hypoteesina eikä siihen päädytä erillisen tutkimuksen avulla (mt. § 55, 77). Toimimisen edellytys ylipäätään on, että joitain asioita ei ollenkaan edes epäillä.

Kielenpuhuja siis tietää kielestä paljon (intuition fenomenologiasta yleisesti ks. Koksvik 2017). Tätä voi pitää tietämisenä ennen muuta *tietää*-verbin etymologisessa merkityksessä (’tuntea tie, seurata jälkiä tai uraa’), ts. kielenpuhuja tavallisesti vain toimii *niin* ja *niin*, intuitiivisesti. Vain erityisissä käytännön tilanteissa (esimerkiksi taivutetaanko *viiveen* vai *viipeen*, jonkun nimen muistaminen, suostuttelutilanne jne.) kieltä koskevaan tietoon tai kielelliseen käytäntöön kohdistetaan itsehavainnoivaa huomiota (ks. luku 5.2.). Myös äidinkielisen kielentutkijan kieltä koskevan kielitajun perusta lepää edellä kuvatussa varmuudessa. Sekin on tavallisen kielenpuhujan kielitajua. Kuitenkin kielitaju on välineenä selvästi tiedostettu ja erikseen harjoitettu metodi osana kielentutkimusta ja sen teoreettista perinnettä. Metodilla pyritään tieteelliseen kuvaukseen osasta kieltä menetelmällä, jota käsiteanalyysin yhtenä tyyppinä voi kuvata **eksplikaationa** eli käsitteellistettävien asioiden yhä selvempänä ilmaisemisena (Itkonen 1978, 301–310; ks. Käsiteanalyysi kielentutkimuksessa tk.; Tieteen termipankki s.v. eksplikaatio).

Kielenpuhujan tiedostamista esiteoreettisista säännöistä lingvistin kielitajun kuvaus eroaa ensiksi selvimmin teoreettisen käsitteistönsä vuoksi. Esimerkissä (6) muut paitsi kielenainesta merkitsevät sanat ovat teoreettisia käsitteitä, mikä etäännyttää tällaisia kuvauksia jo merkittävästi kielenpuhujan huomioista (vrt. esim. 4 ja 5). Näin kielentutkijan esitykset kiinnittyvät kielioppikuvauksen teoreettiseen perinteeseen. Esimerkki (6) on teoreettinen sääntökuvaus. Termit *lekseemi*, *yksikkö* ja *illatiivi* ovat teoreettisia käsitteitä, verbi *taipuu* kuvaa teoreettisesti kuvattavaa operaatiota (sanan ilmiasun muuttumista), ja *suomi* konstituoi yhtenäisen kielen, jonka säännön esitetään kattavan. Tällaiset sääntökuvaukset voidaan laatia lukuisilla eri tavoilla, ja esimerkiksi erilaiset kielioppiteoriat ja -formalismit ovat tästä kuvaava osoitus (ks. Kieliopin tutkimus ja kielioppiteoriat tk.).

(6) Suomessa lekseemi *talo* taipuu yksikön illatiivissa *taloon*. [vrt. 2c, 4]

Toinen ei-kielentutkijan ajattelusta selvästi eroava seikka on se, että kielentutkimuksessa teoreettisen kuvauksen motivaationa on usein koota esimerkin (6) tapaisia sääntöjä yhtenäiseksi ja systemaattiseksi kieliopin järjestelmän esitykseksi tai kuvata yksittäisiä sääntöjä jonkin (osa)järjestelmän valossa. Kolmas seikka koskee tutkijan positiota, joka on tutkimuskohdetta aina jossain määrin olioiva: toisin kuin ei-kielentutkijalla tutkijalla on erityinen havainnoiva suhde kielen käytäntöön ja käyttöinstansseihin yleistämistä ja sen kuvaamista ajatellen. Neljäntenä seikkana voi mainita tutkijan tietoisuuden siitä, että kielitajulla voi operoida vain oman tietoisuuden piirissä ja että se kohdentuu viime kädessä yksilömurteeseen (idiolektiin). Tämä koskee kielitajun kontrollointia ja sen rajoitteiden tiedostamista metodina.

Kielitajun mukaista kieltä voi tutkia myös käyttämällä kolmannen henkilön eli kielenoppaan kielitajua kyselymenetelmillä tai elisitaatiolla (ks. Kysely; Kielten dokumentointi ja kieliopin kuvaus tk.). Tämä on luonnollista silloin, kun tutkijan oma kielitaju ei tunnu metodisesti enää riittävän ja kun halutaan erityisesti tietoa muiden kielitajusta.

Herää helposti kysymys, kuinka paljon esimerkkien (2–6) tapaisia selviä tapauksia on. Arvellen selviä tapauksia on yksilöitävissä tuhansia, tai jos huomioidaan sanojen merkitys, niin kymmeniä tuhansia. Paljon riippuu itse sääntökuvauksenkin tarkkuudesta ja abstraktiotasosta. Kielentutkimuksen kannalta nousee myös kysymys siitä, miten tarkan kuvauksen voi vielä ylipäätään olettaa vastaavan varmuutta tietoisella tasolla.[9] Esimerkkeinä sääntökokoelman systematisaatioista eli kielioppikuvauksista, joissa metodina on käytetty kielitajua, edustavat suomen osalta esimerkiksi morfo(fono)logiassa Karlsson (1982b) ja syntaksissa Hakulinen ja Karlsson (1995 [1979]).

Vielä on tarkennettava selvien tapausten varmuuden lähdettä. Kielitaju rajautuu niin ei-kielentutkijalla kuin kielentutkijallakin omaan mieleen, ja koettu varmuus on aina viime kädessä subjektiivista. Wittgenstein (1999 [1969], § 284, 487, 634) esittää, että kielenpuhuja ei kykene tarjoamaan loogisesti riippumatonta perustetta varmuudelleen, josta hän voi kuitenkin subjektiivisessa erehtymättömyyden tunteessaan tiedostua. Selvät tapaukset kielessä ja niiden muodostama kokonaisuus vain *on*, tai paremminkin *niin* ja *niin* vain perustavasti *tapahtuu*, kun kieltä tulee käyttäneeksi. Kielentutkimuksessa tätä perustaa pyritään saattamaan teoreettiseksi kuvaukseksi.

Taustalla ovat säännöt, normit ja konventiot (”näin *pitää* toimia”), jotka ovat yliyksilöllisiä eli sosiaalisia, kieliyhteisön kaikkien jäsenien tai ainakin enemmistön kesken pääosin jaettuja (esim. Itkonen 2003, luku 19). Kielenpuhujan varmuuden takaavat siis normit, jotka vallitsevat, kohdistavat velvoittavuutta yksilön toimintaan tämän ehkä huomaamatta ja tarpeen vaatiessa käyvät epäsuorasti ilmi, mikäli normista livetään liiaksi. Keskeisin kollektiivisesti jaetun kielellisen normin peruste on ns. yksityiskieli-argumentti (engl. *private language argument*, ks. Itkonen 2003, luku 21). Yksilö ei voi kaikessa vakuuttuneisuudessaankaan tukeutua kieleen, jota vain hän yksin ymmärtää. Jo jokin esimerkin (1c) tapainen kielenkäyttö johtaa helposti ymmärtämättömyyteen, mikäli ympäristö ei osaa, tunne tai tunnista leikkikielen koodia (vrt. kuviossa 1 alue ei-p).

Vastaavasti kielentutkijan tiedollinen peruste kielitajunsa käytölle on, että selvät tapaukset ovat normatiivisia. Kielitaju on yhtä subjektiivinen

kuin ei-kielentutkijalla ja metodina periaatteessa toistamaton (replikoimaton) siinä mielessä, että se ei ole ilmipantava ja selvästi ositettavissa oleva proseduuri, jonka voi mallin mukaan toistaa. Sen sijaan sen pätevyys mittautetaan saattamalla kielitajulla saatu kielenaines ja sen varassa tehdyt teoreettiset päätelmät ja esitykset muiden kielentutkijoiden arvioitavaksi. Jossain määrin tätä tekevät tavalliset kielenpuhujatkin, mutta kielentutkimuksessa se on olennainen osa itse metodin käyttöä.

## 3. Terminologiasta ja lähikäsitteistä

Kuten on jo käynyt ilmi, kielitajua koskevia tai sitä lähellä olevia käsitteitä on lukuisia, ja aina terminologisia eroja ei tehdä riittävän selviksi (ks. tietolaatikko 1). Tässä luvussa selostetaan näitä eri käsitteiden yhteneväisyyksiä ja eroja.

Termillä *introspektio* (lat. *intrōspectus* < *intrōspicere* ’katsoa, kurkistaa sisään’, kuv. ’luoda tutkiva katse’) tarkoitetaan yleisesti ensimmäisen persoonan tietoisen mielen subjektiivisten tajunnansisältöjen tarkastelua ja arviointia. Tajunnansisällöt voivat olla esimerkiksi tunnetiloja, muistoja, mentaalisia prosesseja tai propositionaalisia tietosisältöjä.[10] Näin introspektio voi olla sekä passiivista (kokevaa) että aktiivista (tietoisesti kohdistuvaa). Kielitieteessä termillä on myös hieman toisenlainen käyttö: sillä tarkoitetaan kielellisten ilmausten arviointia mm. niiden hyväksyttävyyden tai merkityksen kannalta.

Tieteen termipankin antamat määritelmät *introspektiolle* vaihtelevat nekin tieteenaloittain. Kielitieteellisen määritelmän mukaan *introspektio* tarkoittaa sitä, että ”tutkija nojaa omaan kielitaitoonsa ja arviointikykyynsä tutkiessaan kieltä” (Kielitiede → *itsehavainnointi* | *introspektio*, ks. myös Karlsson 1998, 21). Filosofinen määritelmä taas kuuluu seuraavasti: ”oman tietoisuuden sisältöjen, tilojen ja rakenteiden havainnointi” (Filosofia → *introspektio*). Filosofisessa määritelmässä *havainnoinnilla* tarkoitetaan nimenomaan itsereflektiota eli havainnoijan oman tietoisuuden kohdistamista havainnoijan omiin tajunnansisältöihin. Tällöin

havainnointitapahtuma ja sen osat sijoittuvat kokonaisuudessaan tietoisuuteen eli arkisesti ilmaisten ”pään sisään”. Tämä on syytä pitää erossa aisteilla saaduista havainnoista, joiden kohteet sijaitsevat tyypillisesti havaitsijan kognition ulkopuolella, aistittavassa (ulko)maailmassa.

Toisinaan *introspektiolla* siis tarkoitetaan kielentutkimuksen metodina itsehavainnointia, jolla tutkija kykenee tietoisesti tekemään selkoa äidinkielekseen sisäistämänsä kielijärjestelmän normistosta ja kielenkäyttöä koskevista konventioista. On ehkä syytä erikseen korostaa, että *normistolla* ja *konventioilla* ei tässä tarkoiteta kielenohjailua ja kielenhuoltoa ja sen kohteena olevaa yleiskieltä vaan kielenkäyttäjän sisäistä tietoa, joka voi koskea yhtä hyvin normittamatonta kielimuotoa, kuten murretta tai sosiolektia.

*Intuitiolle* (lat. *intuitio* ’näkemys’ < *intueri* ’katsoa, pitää jonakin’) on annettu myös hieman eri merkityksiä eri vahvuusasteina. Yhden tulkinnan mukaan se on välitöntä non-propositionaalista varmuutta, että pitää olla (esim. *korven*) tai ei pidä olla (esim. **korpin* kuten 2c:ssä) (Rorty 1967, 204). Intuition voi ajatella kontrolloimattomaksi ja esitietoiseksi tasoksi varmuutta, ja vain introspektiolla tämä voidaan saattaa (a)teoreettisina kuvauksina ilmi. Suhteessa metodiin introspektio on ”varmuuden jälkeisenä” ja sikäli retrospektiivisenä tavallaan metodisempi kuin sitä perustavampi intuitio. (Ks. Devitt 2006, 491; Raami 2015.)

Kuten todettua **kielitajulla** tarkoitetaan tiedostettua ja tiedostamatonta omaa äidinkieltä koskevaa ymmärrystä (intuitiosta tietoisena päättelynä ks. Koksvik 2013). Se kattaa tavallisesti sekä intuition että introspektion, mikäli näiden käsitteiden kesken ei ole tehty eroa (vrt. tietolaatikko 1). **Kielivaisto** (myös kielikyky, kompetenssi) viittaa enemmän kielentämättömään intuitioon, jota voidaan saattaa tietoisen introspektion avulla joltain osin tietoisesti tarkasteltavaksi (ks. Karlsson 1998, 22). **Kielitaidolla** tarkoitetaan tyypillisesti sekä oman kielen käytännöllistä kielenkäyttötaitoa että vieraiden kielten osaamista käytännössä. **Kielitieto** koskee hankittavissa olevia erilaisia formaaleja tietoja kielestä. Samoin **kielen** (ja kulttuurin) **tuntemus** on käytännön kielitaidon kerryttämää jotain kieltä ja kieliyhteisöä koskevaa kulttuurista tietämystä. **Kielitietoisuudella** taas viitataan tavallisesti ideologiseen ja kielelliseen vallankäyttöön ja vaikuttamiseen liittyviin ilmiöihin sekä laajasti kielen erilaisen

monimuotoisuuden tiedostamiseen yhteiskunnassa (monikielisyys, monilukutaito yms.) (ks. myös Piippo, Vaattovaara & Voutilainen 2016). *Kielitietoisuuden* käsitettä on sovellettu myös kansanlingvistiikassa tutkittaessa kielenkäyttäjien kielitajua (esim. Preston 1996). Kielitieto, -taito, -tietoisuus ja kielen tuntemus ovat yleensä opetuksen ja koulutuksen myötä hankittavissa, ja kielentutkijan kielitajun tukena ne ovat usein hyödyksi. Esimerkiksi tieto jonkin yksittäiskielen yleisrakenteesta, kuvausperinteestä ja muista teoreettisista taustoista on elimellinen osa tutkimusta, mutta ne ovat kuitenkin selvästi eri asia kuin kielitaju metodina.

Toisinaan on osin leikkisästi ja osin perustellusti arvostellenkin käytetty kielitajusta metodina nimitystä *nojatuolilingvistiikka* ja kutsuttu esimerkiksi itse aineistoksi tuotettuja lauseita *nojatuoli*- tai *laboratoriolauseiksi* (ks. Fillmore 1992, 35; Huumo 2007). Kritiikki on koskenut muun muassa sellaisia keksittyjä lause-esimerkkejä, joita ei löydy aidosta kielenkäytöstä mutta jotka periaatteessa ovat kielenmukaisia (ks. luku 4.1.). Vähemmän perusteltua on ollut kritiikki, joka kieltää kokonaan kielen käyttöä koskevilla esimerkeillä operoimisen kielitajun varassa. Yleisemminkin puhutaan nojatuolifilosofoinnista tai -teoretisoinnista sellaisen tutkimuksen yhteydessä, jossa esimerkiksi ei kerätä ulkoista aineistoa tutkimuksen pohjaksi. Keskeisin lähde tällaiselle lisäepiteetille on se kritiikki, jota kielitajumetodin yksinomaista käyttöä kohtaan on esitetty kielentutkimuksen nk. empiirisen käänteen jälkeen (ks. luku 4.2.). (Kielentutkimuksen käänteistä ks. Hamunen, Nieminen, Kelomäki & Dufva 2019.)

**Tietolaatikko 1. *Intuitio*, *introspektio* vai *kielitaju*?**

Tutkijan oman kielitajun käyttöä lingvistisen analyysin metodina on kutsuttu vaihtelevasti *introspektioksi* ja *intuitioksi*. Itkonen ja Pajunen (2010, 10–12) esittävät eron seuraavasti:

- *Introspektio*: tietoinen prosessi, jonka kohteena ovat (yksilölliset) kokemukset tai tietoisuuden sisällöt, esimerkiksi mielikuvat ja tunnesävyt.

- *Intuitio*: tiedonhankinta-akti, joka kohdistuu luonteeltaan sosiaaliseen normiin.

Koska intuitio kohdistuu normiin, vain sen avulla kielentutkija voi arvioida, onko ilmaus korrekti vai epäkorrekti (**talo punainen tämä*). Näin määriteltynä tässä artikkelissa käytetty termi *kielitaju* tarkoittaa siis samaa kuin kieltä koskeva *intuitio*.

# 4. Kielitajun tutkimushistoriaa

## 4.1. Kielitaju kielentutkimuksen historiassa

Metodeja koskeva metateoreettinen tietoisuus on kielentutkimuksessa verraten nuorta. Tutkimusta koskeviin metodologisiin seikkoihin on kohdistettu eksplisiittistä huomiota vasta suhteellisen lyhyen aikaa, vaikka esimerkiksi länsimaisen kielentutkimuksen historia palaa monessa suhteessa antiikkiin asti (esim. Robins 1969 [1967], luku 1–2; Timofeeva 2006). Kuitenkin perinteisissä kieliopeissa tapana on ollut rakentaa kielen kuvaus joko kokonaan tai merkittäviltä osin kuvaajan itse keksimien esimerkkien eli kielitajun varaan. Esimerkiksi Apollonios Dyskoloksen teoksessa *Syntaksista* (*Peri syntakseōs*) esimerkeistä noin neljännes on grammaatikon itsensä tuottamia lauseita. Loput on poimittu antiikin klassikoista, erityisesti Homerokselta. (Householder 1981, 5.) Apollonios myös ainakin kerran mainitsee, että joidenkin lauseiden virheellisyyttä

(esim. **grapso an* 'aion kirjoittaa') ei voi perustella muuten kuin vetoamalla välittömään intuitioon (kr. *antilēpsis*) (Apollonius 1981, 161).[11] Virheellisyyden tai epäkieliopillisuuden lajeja hän esittää kahta tyyppiä: syntaktisia virheitä (kr. *soloikismos*) ja sanakohtaisia artikulointivirheitä (kr. *barbarismos*) (Apollonius 1981, 7, 155–162). Edelliset vastaavat kielitajulla tavoitettavia luonnollisia normeja, jälkimmäiset preskriptiivisiä (oikeakielisyys)normeja. Yleisesti grammaattisilla kuvauksilla ohjattiin käyttämään kieltä oikein ja jonkin esteettisen mittapuun mukaan, esimerkiksi hyvien kirjailijoiden tai sanankäyttäjien mallin mukaisesti.

Intuitioon perustuvan kieliopillisuuden ja introspektiolla tuotetun kielenaineksen varassa argumentointi on (länsimaisen) kielioppitradition ensiaskelista ollut osa kuvauksen käytännön metodiikkaa. Metodologia itsessään ei kuitenkaan pitkään aikaan noussut näkyviin. Tälle lienee lukuisia syitä: 1) erilaisia metodeja ei yksinkertaisesti ollut, ja olemassa olevaa käytäntöä – tekstilähteiden käyttöä ja kielitajua – ei tarkasteltu metodologian kannalta; 2) kuvauksessa jouduttiin aluksi keskittymään peruskäsitteiden ja muiden kuvauskäytänteiden luomiseen; 3) huomio kohdistui itse deskriptioon ja usein siihenkin, mitä kielen oletettiin paljastavan (esim. ajattelun universaalit kategoriat, ihmisen alkukieli, kansansielu tms.).

Metodologiset huomiot yleisesti nousivat selvemmin esiin vasta 1800-luvun jälkipuoliskolla nuorgrammaatikkojen historiallis-vertailevaa menetelmää kohtaan esittämästä kritiikistä, jonka keskiössä olivat muun muassa positivismin hengessä vaateet empiirisestä aineistosta, objektiivisuudesta ja muun muassa psykologismin ja kaikenlaisen käsitteellisen spekuloinnin välttämisestä (ks. Kelomäki 2014, 264; varhaisesta anti-psykologismista ks. Kantor 1952 [1936], 46–56; Carroll 1955, 71–82). Referenssitieteenä olivat luonnontieteet, ja jo yksinään tällöisiä metodologisia huomioita sinänsä voi pitää merkittävinä kielitiedettä modernisoivina tekijöinä. Kuitenkaan psykologisointien välttämisen vuoksi kielitajua ei löydy nuorgrammatiikan metodologiasta. Kielentutkimuksen ja psykologian yleisestä suhteesta 1800- ja 1900-luvun molemmin puolin tarjoaa tiiviin yleiskuvan Kess (1983).

Karkeasti yleistäen erityisesti kahdella tutkijalla oli keskeinen vaikutus kielitajun esiinnousulle kielentutkimuksessa, Ferdinand de

Saussurella ja Noam Chomskylla. De Saussure esitti *Yleisen kielitieteen kurssissaan* (2014 [1916], 81–95) käsitteellisen jaon kielen järjestelmään ja puhuntaan (ransk. *langue*/*parole*). Lingvistiikan ensisijaisen kohteen hän siirsi järjestelmän hahmottamiseen. Tämä loi mahdollisuuden sille, että oli ylipäätään olemassa jokin käsitteellinen kielisysteemi, joka on omaksuttavissa, periaatteessa tutkittavissa ja erillään käytännön puhunnasta. Historiallisen kielitieteen tutkimuskohde, kielenmuutokset, ei ollut periaatteessakaan kielitajulla tavoitettavissa. De Saussure kuitenkin sijoittaa kielijärjestelmän sosiaaliselle tasolle ja toteaa, että ”[k]ieli ei ole kielenpuhujan kyky, vaan tuote, jonka yksilö omaksuu passiivisesti” (de Saussure 2014 [1916], 88). Kieli on siis ensi sijassa sosiaalinen instituutio, ja sen yksilöllinen omaksuma on toissijainen.

Kuitenkin de Saussure myös toteaa, että erilaisten kielen edellyttämien fysiologisten seikkojen ja elinten (puhe-elimet, artikulaatiorutiinit, aivot) ”yläpuolella” on yleisempi merkkien hallinnasta vastaava kielikyky (de Saussure 2014 [1916], 84). Tämä on tulkittavissa intuitioksi, vaikka de Saussure ei tätä termiä käytäkään. Myöskään *introspektiota* ei terminä mainita, mutta sen sijaan tehdään ero *subjektiivisen* ja *objektiivisen analyysin* välillä (de Saussure 2014 [1916], 305). Edellinen tarkoittaa kielenpuhujien harjoittamaa kielen yksiköiden alinomaista analyysia, jälkimmäinen diakronista tutkimusta. Selostus (de Saussure 2014 [1916], 305–312) kielenpuhujan subjektiivisesta analyysista luo vaikutelman, että kyse on tavallisesta kielenpuhujasta ja tämän (online-)kielitajusta (ks. luku 5.2.). De Saussure ei esitä eksplisiittistä kielitajumetodia.

Koska de Saussuren *Kurssin* perimmäinen tavoite oli asettaa yleisellä tasolla teoreettis-käsitteellinen perusta modernille (synkroniselle) kielentutkimukselle, siitä ei löydy metodista opastusta tai metodologiaa koskevaa tietoteoreettista pohdiskelua juuri edellä kuvattua enempää. Edellisen varaan syntyi kuitenkin useita strukturalistisia kielentutkimuksen suuntauksia (ks. esim. Helbig 1973, 342), joissa on eri tavoin otettu suoraan tai epäsuorasti kantaa kielitajuun. Karkean kolmijaon strukturalististen suuntausten ja tutkijoiden kesken voi tehdä epistemologisten kannanottojen perusteella seuraavasti: 1) pelkästään kielijärjestelmän kuvaukseen keskittyvät eli ei-kantaaottavat (platonistit), 2) anti-psykologiset (mekanistit) ja 3) kielitajun huomioivat (mentalistit).

Näistä kaksi jälkimmäistä ovat tässä huomionarvoisia. Bloomfieldin (1957 [1933], 32–41, 139–142) behaviorismiin pohjaava mekanistinen käsitys on tunnettu siitä, että sen mukaan kieltä tuli ensi sijassa tutkia käytössä ja vain havaittavan toiminnan tutkimiseen soveltuvin menetelmin ja että merkitykseen ei kielentutkimuksen silloisin menetelmin päässy käsiksi (vrt. Bloomfield 1983 [1914], luku 3; ks. myös Fries 1970, 203–217).[12] Käytännössä kuitenkin esimerkiksi Bloomfieldin (1957 [1933]) englantia kuvaavat esimerkit olivat kirjoittajansa kielitajun varaisia.

Varsinaisia mentalisteja ei juurikaan ollut ennen Chomskya, mutta paikka paikoin tapaa pohdintaa kielitajun käytöstä metodina. Tesnière (1959, luku 18) esimerkiksi esittää, että kielitajua on vallitsevien menetelmien vastaisena ja intuitioon nojaavana moitittu liian subjektiiviseksi. Hän toteaa kuitenkin seuraavaa:[13]

> Tämä pelko on vielä helpommin kumottavissa. On totta, että introspektio [kielitaju] nojaa intuitioon. Mutta se nojaa myös sisäiseen kokemukseen. Siksi se on kokemusperäinen [*expérimentale*] menetelmä ja siten objektiivinen. (Mts. 38.)
>
> Introspektio [Kielitaju] on objektiivista myös siitä syystä, että se koskee tosiasioita [*faits*]. Nämä tosiasiat ovat toki abstrakteja. Mutta ei ole olemassa pelkästään konkreettisia tosiasioita. Voidaan jopa sanoa, että syntaksissa kaikki tosiasiat ovat abstrakteja, koska loppujen lopuksi ainoastaan niiden merkitsin [*marquant*] on konkreettinen. On siis tuskin muuta menetelmää, jolla voidaan saavuttaa täysin abstraktit syntaksiin liittyvät tosiasiat kuin sisäinen analyysi. (Mp.)

Silti Tesnière (mp.) myöntää, että kielitajua on käytettävä varoen: tutkijan on oltava äidinkielinen tutkittavan kielen puhuja, kielitaju ei ole sisäsyntyistä vaan erikseen harjaannutettava taito, ja sitä voi soveltaa syntaksissa sanojen keskinäisiä suhteita tutkittaessa (esim. dependenssi).

Myös jotkut varhaiset ei-strukturalistiset tutkijat saattoivat ottaa kantaa kielitajun käyttöön ja olemassaoloon, vaikka heidän kiinnostuksensa kieltä kohtaan virisi yleensä nimenomaan kielen kontekstuaalisuuden ja käytön (*parole*) tutkimisesta. Esimerkiksi Gardiner (1951 [1932], 343–344)

esittää, että havaittavia maailman tapahtumia tehdään ymmärrettäviksi tietoisen järkeilyn ja introspektiivisen tiedon avulla. Yleisemmin ei-strukturalistit kuitenkin vieroksuivat yksinomaista kielitajun käyttöä, mikä juontui esimerkiksi (kieli)psykologian piirissä mentalismiin kohdistuneesta kritiikistä ja behaviorismin myötä erityisestä kiinnostuksesta kielelliseen käyttäytymiseen (ks. Carroll 1955, 75). Esimerkiksi John R. Firth (1957, 175–176, 179) pitää kieltä kaikkiaan kontekstuaalisena. Se ilmenee ja siihen on pääsy vain tekstien kautta. Siksi kielitaju ei ole riittävä metodi.

Vielä Zellig Harris (1969 [1951], 12–13) piti amerikkalaisen strukturalismin kielenkuvauksen kohteena kerätyn korpusmateriaalin analysointia. Kuten todettua Chomsky (2001 [1965], 3–27) kuitenkin ohjelmallisesti siirsi kielentutkimuksen painopistettä voimakkaasti mentalismin puolelle. Generatiivinen kielioppi oli ensimmäinen kokonaisvaltainen synkroninen kielioppiteoria ja erään tulkinnan mukaan samalla myös ensimmäinen strukturalismin kognitiivinen versio ja väljässä merkityksessä kognitiivisen kielentutkimuksen aloittaja.

Chomskylla on oma kielitajua sivuava terminologiansa. Hänen mukaansa *kompetenssi* on kielenpuhujan sisäistämää implisiittistä tietoa (engl. *intuition, tacit knowledge*) kielestä ja erityisesti sen lauseista. Chomskyn ja hänen seuraajiensa mielenkiinto kohdistuikin ennen kaikkea abstraktiin syntaksiin, joka ilmentää kompetenssia. Kompetenssiin ei tutkijalla kuitenkaan ole suoraa pääsyä, sillä sen perusta on chomskylaisen ajattelun mukaan biologisperäisessä synnynnäisyydessä, joka on tietoisuuden alista. Vaikka yksilöstä tai ideaalista puhuja-kuulijasta siis tulikin biologisen kompetenssin kantaja, oli kielen lauseisiin silti suorin metodinen pääsy tutkijan oman kielitajun varassa. *Performanssi* sen sijaan on kieltä aidoissa käyttöyhteyksissään. Chomsky kuitenkin sivuutti sen siksi, että puheen erilaiset toistot, katkot, korjaukset, epäröinnit ja muut häiriötekijät peittävät hänen mukaansa alleen kieliopin ja kompetenssin säännönmukaisuutta. (Chomsky 2001 [1965].) Silti kielitajuakin on pidettävä juuri yhtenä performanssin ilmentymänä.

Generatiiviselle syntaksille on ollut erityisen leimallista tutkijan oman kielitajun varassa tuottaa erilaisia syntaksin abstrakteja, analyysilla esiin tuotavia ilmiöitä edustavia lause-esimerkkejä. Esimerkit (7a) ja (7b)

näyttävät **pintarakenteeltaan** samanlaisilta. Ne ovat kuitenkin **syvärakenteeltaan** erilaisia, sillä transitiivinen infinitiivitäydennys, jota muunnellaan aktiivissa ja passiivissa, saa erilaisia tulkintoja englannin *persuade*- ja *expect*-verbeissä. Esimerkit (9a) ja (9b) ovat totuusarvoltaan samoja: mikäli (9a) on totta, myös (9b) on. Sama ei päde esimerkkeihin (8a) ja (8b); niiden totuusarvot eivät ole toisistaan riippuvaisia. Chomskyn mukaan kieliopin teoreettisen kuvauksen tulisi kohdistua nimenomaan tällaiseen kompetenssin piilevään tietoon.

(7a) *I persuaded John to leave*
(7b) *I expected John to leave*
(8a) *I persuaded a specialist to examine John*
(8b) *I persuaded John to be examined by a specialist*
(9a) *I expected a specialist to examine John*
(9b) *I expected John to be examined by a specialist* (Chomsky 2001 [1965], 22).

Syntaktisen tiedon abstraktistavan pinta- ja syvärakenteen eron lisäksi toinen merkittävä käsitepari Chomskylla (2001 [1965], 10–15) on ollut *kieliopillisuus* (engl. *grammaticalness*) ja *hyväksyttävyys* (engl. *acceptability*). Edellinen koskee kompetenssia, jälkimmäinen performanssia. Esimerkki (10a) on kieliopillinen eli englannin kieliopin mukainen. Vastaavasti (10b) ei ole, sillä oikealle haarautuvissa rakenteissa relatiivipronomini korreloi yleensä välittömästi edeltävän sanan kanssa, ei rakenteiden yli. Esimerkin (10a) toistuva *that*-rakenne ilmentää kompetenssin ja kieliopin kykyä tuottaa periaatteessa määrättömän pitkiä kieliopillisia rakenteita (rekursiivisuus eli toistettavuus). Tällaiset kielen ilmiöt ovat Chomskyn (2001 [1965], 15–16) mukaan todiste kielikyvyn rajaamattomuudesta. Vastaavasti molemmat esimerkit (11a) ja (11b) ovat kieliopillisia, mutta (11a) on käytössään hyväksyttävämpi kuin (11b). Se on luonnollisempi, helpommin ja nopeammin ymmärrettävissä sekä vähemmän kömpelö kuin (11b). Esimerkki (12) kuvaa erityisen ei-hyväksyttävää mutta kieliopillista englannin lausetta.

(10a) *This is the cat that caught the rat that stole the cheese*
(10b) **This is the* ***cat***$_i$ *that caught the* ***rat***$_j$ *that stole the cheese* ***that***$_i$ *eat* ***it***$_j$
(11a) *I called* ***up*** *the man who wrote the book that you told me about*
(11b) *I called the man who wrote the book that you told me about* ***up***
(12) *the man who the boy who the students recognized pointed out is a friend of mine* (Chomsky 2001[1965], 10–13).

Dąbrowska (1996) tutki äidinkielisten kielenpuhujien kykyä ymmärtää kieliopillisia mutta hyväksyttävyydeltään vaihtelevia englannin lauseita poimimalla generatiivisista tutkimuksista esimerkkien (13–15) kaltaisia tapauksia arvioitavaksi. Tutkittavilta kysyttiin muun muassa, kuka kuvatussa tilanteessa tekee kelle ja mitä. Tulos oli puhutteleva: kyky ymmärtää lauseita korreloi vastaajien koulutustaustan kanssa, vaikka kielitajun takaisen kompetenssin pitäisi periaatteessa olla yhtäläinen kaikille kielenpuhujille (ks. myös Spencer 1973). Vastaajina oli muun muassa yliopiston siivoojia, lehtoreita ja opiskelijoita, joista kukaan ei ollut kielentutkija. Myös koulutetuille vastaajille lauseiden tulkitseminen oli haastavaa, ei oletuksen mukaisesti automaattista. Kognitiivisten tulkintarutiinien puuttuessa tällaisten esimerkkien käyttäminen sinänsä uskottavista teoreettisista kieliopin ilmiöistä (esim. jonkin rakenteen toistuvuus) vetääkin helposti vesiperän.

Esimerkit (13–15) asettuvat kuvion 1 (ks. luku 2.1.) alueelle mahdollisesti p. Ne ovat kieliopillisia, mutta koska niiden kaltaisia löytyy hyvin harvoin todellisesta kielenkäytöstä, joutuu vakavasti kysymään, mikä niiden arvo on edes teoreettisintakaan kielikykyä ajatellen. Luottamus kielitajulla tuotettujen esimerkkien käyttöön on kokenut ehkä suurimman arvonalennuksensa juuri tällaisilla lauseilla operoinnin vuoksi. Tällaisia lauseita on käytetty erittäin paljon kielitieteessä juuri generatiivisen teorian piirissä. Schütze (1996) on perusteellinen selonteko tällaisen metodologian teoreettisista hankaluuksista (ks. myös Wasow & Arnold 2005, Chomskyn varhaisista metodologia-ajatuksista Karlsson 2008).

(13) *The manager knew that the fact that taking good care of herself was essential upset Alice.*

(14) *That dog will be difficult to persuade the little girl to play with.*
(15) *John thought that the fact that criticizing himself was hard surprised Mary.*

Tämän jälkeen 1970-luvun kuluessa ja sen jälkeenkin erilaisia, aluksi erityisesti syntaksiin keskittyviä kielioppimalleja syntyi lukuisia. Näitä esittelevät tiiviisti muun muassa Moravcsik ja Wirth (1980) sekä Droste ja Joseph (1991). Kielitaju on ollut aina keskeinen kuvattavien kielenilmiöiden hankintatapa erilaisissa teoreettisissa kielenkuvausmalleissa. Myös monien nykyään olemassa olevien mallien perusta on alkuaan rakennettu kielitajulla tuotettujen esimerkkien varaan. Tämä näkyy joko käytetyistä esimerkeistä (esim. Langacker 1987) tai sekä esimerkeistä että eksplisiittisestä kielitajun asemoimisesta kielentutkimuksen metodologiassa (esim. Talmy 2003 [2000], 4–6).

## 4.2. Kielitajulle esitettyjä rajoituksia ja kritiikkiä

Kielitajun käyttöä on jonkin verran kritisoitu. Motivaationa on ollut erityisesti generatiivisessa suuntauksessa vallinnut mentalismi, joka on ilmennyt muun muassa tutkijan oman kielitajun melkein yksinomaisena käyttönä. Edellä esittelimme jo Dąbrowskan (1996) tutkimuksia yhtenä esimerkkinä kielitajun käytön kritiikistä.

Myös toinen generatiivisen suuntauksen kritiikki juontuu metodista. Esitykset universaalikieliopista ovat nojanneet viime kädessä englanninkielisten tutkijoiden oman kielitajunsa varassa muodostamiin esimerkkeihin. Tiedetään kuitenkin, että yhden kielen sisällä tietyn valikoituneen ihmisryhmän (akateemiset kielentutkijat) kielitajun varaisen aineksen ei voi ajatella edustavan edes yhtä kieltä, saati koko ihmiskieltä. Lisäksi taustalla olevan kielitieteellisen teoretisoinnin voi jo nähdä vaikuttavan ohjaavasti esimerkkityyppien sepittämiseenkin (esim. 13–15). Myös käyttäytymistieteellisessä tutkimuksessa tiedetään, että muun muassa havaintojen tekeminen (esim. kielestä) ja niiden selostaminen (esim. kielitajulla) on vahvasti kulttuuri-, koulutus- ja muihin taustatekijöihin sidoksissa (ks. Henrich, Heine & Norenzayan 2010). Näin

ollen käytettävän aineiston, sen yleistettävyyden ja selittämisen suhdetta tulee kielentutkimuksessakin tarkkaan pohtia metodista riippumatta.

Kun erotetaan kielitajun kritiikistä erikseen nimenomaan generatiivisesta perinteestä motivoituva ja siihen kohdistuva kritiikki, voidaan esitetyt varaukset jakaa karkeasti kahtia 1) kielitajun yleistä olemusta koskeviin ja 2) kielitajun varassa saatavaa tietoa koskeviin. Ensiksi kielitajun olemusta koskevat samanlaiset ominaisuudet kuin kaikkea mentaalista toimintaa: siihen on viime kädessä pääsy vain yksilöllä itsellään. Se on solipsistista. Kielitajun tuotokset saatetaan kuitenkin tutkimuksessa kollektiivisesti arvioitaviksi osana lingvististä argumentaatiota. Lisäksi tyypilliset selvät tapaukset perustuvat yhteisiin sosiaalisiin normeihin, jotka tutkijakin on sisäistänyt kieliyhteisön jäsenenä ja joiden kielenmukaisuutta ei tavallisesti ole tarpeen erikseen oikeuttaa (esim. korpusaineistolla). Sama vasta-argumentointi käy myös väitteisiin siitä, että kielitaju koskisi aina vain yksilön idiolektia tai että kielitaju olisi aina epävarmaa ja vaihtelevaa. Selvissä tapauksissa näin ei ole. Epäselvissä tai rajatapauksissa asiaan saadaan yleensä selkoa vertailemalla eri puhujien kielitajuja sekä käyttämällä muita metodeja (esim. korpus- tai kyselytutkimus).

Toinen kielitajun olemukseen liittyvä seikka on äidinkielisen puhujan vaatimus. Silti myös esimerkiksi vierasta kieltä tutkiva kenttälingvisti käyttää ko. kieltä koskevaa ja alati kehittyvää kielitajuaan osana analyysivälineistöä. Tämä tapahtuu ainakin niin, että kielenmukaisuudet erotetaan epäkielenmukaisuuksista ja aidot aineistoesimerkit luokitellaan esimerkiksi kielioppikuvauksen kategorioiden mukaisesti.

Niinpä vaikka kielitaju onkin yksilökohtainen ja yleensä vahvimmillaan äidinkielinen kyky, sillä tuotetut esimerkit kohdistuvat yleensä sosiaalisesti jaettuun ja ylipäätään jaettavissa olevaan tietoon, jonka varassa esitetyt väitteet asetetaan kriittisesti tarkasteltaviksi. Silti on tärkeää, että tutkija tiedostaa kielitajunsa tietyn yksilökohtaisuuden, joka ilmenee yleensä viimeistään kuvion 1 alueiden p ja mahdollisesti p rajankäyntinä (ks. luku 2.1.). Epävarmoissa tilanteissa on aina käytännöllistä hakea tukea muista aineistonhankintametodeista. Erillisen tutkimuksen aihe olisi selvittää, miten kielitaju toimii taustatekijänä tutkijalle vieraita kieliä tutkittaessa.

Toiseksi kielitajulla saatavaan tietoon liittyy erityisesti myös tärkeitä metodiin itseensä kohdistuvia tutkimusaiheita. Toisin sanoen osan esitetystä kritiikistä tai varauksista voi kääntää aidoksi tutkimuskysymykseksi kielitajusta metodina. Yksi tällainen on varauksellisesti esitetty kysymys siitä, millainen on nk. selvien tapausten kaanon. Millainen kokonaisuus niistä mahdollisesti muodostuu? On selvää, että kielitajut eivät ole kaikkinensa identtiset yksilöiden kesken. Sen sijaan selvien tapausten osalta voi kysyä, tarkoitetaanko niillä selviä yksittäistapauksia (*korven* 'korpi-gen' on suomea) vai tyyppitasolle abstrahoituja ja sellaisina kuvattavia ilmiöitä (*korven* edustaa suomen *i* : *e* -nominien taivutustyyppiä).[14] Edellisiä on paljon ja joskus määrättömästi (esim. lause-esimerkeille ei ole ylärajaa), jälkimmäinen on kiinteästi kytköksissä kuvauksen teoriaan.

Toinen kysymys on kielitajuesimerkkien luonne aineistona. Esimerkit ovat niin kiinteä osa jonkin ilmiön teoreettista kuvausta, että esimerkiksi Lehmann (2004, 293–294) taipuu pitämään kielitajuesimerkistöä muuna kuin (aitona korpus)aineistona, jona hän pitää lähinnä empiiristä, havainnoitua dataa. Silti esimerkit ovat usein tarkoitukseensa käypiä, joten niiden episteemistä luonnetta ja asemaa on hyvä pohtia (ks. myös luku 5.1.).

Kolmas kysymys on kielitajuesimerkkien tyypillinen kirjoitetun yleiskielen mukaisuus. Tämä johtaa kysymään, koskeeko kielitaju aina vain kirjoitettua yleiskieltä, joka on lopulta melko rajattu ja keinotekoinen kielen varieteetti. Näin voi joskus olla. Silti kuvattu ilmiö saattaa olla niin laaja-alainenkin, että se koskee varmuudella koko kieltä (esim. suomen sijat ovat suffikseja). Tällöin yleiskielinen kirjoitusmuoto toimii tavallaan vain havainnollistavana välineenä koko kieltä koskevan tyyppitason ilmiölle.

Vaikka tutkimuksen paino olisikin kielitajun käytössä, voi olla varmuuden vuoksi hyvä kerätä ilmiötä havainnollistavia näytteitä myös kielen muista, puhutuista tai kirjoitetuista muodoista. Yleisesti kirjoitettu esitystapa liittyy kielentutkimukselle laajemmin esitettyyn kritiikkiin kirjoitetun kielen harhasta tai skriptismistä (engl. *written language bias,* ks. Linell 2005; *scriptism,* ks. R. Harris 2009). Kritiikin keskiössä on se kehämäisyys, jolla puhuttu kieli saatetaan erilaisin kirjoitusmerkein

(grafeemit, litteraatiot, transkribointi yms.) symbolimuotoon ja näistä tehdyt päätelmät palautetaan kielen ominaisuuksiksi. Näin kielestä annetaan kritiikin mukaan liian oliollistettu kuva.

Kielitajua on kritisoitu myös siitä, että sillä ei voi kuvata teksti- tai keskustelutason ilmiöitä. Kuitenkin kielitajun varassa on pääsy moniin tämänkin tason seikkoihin (ks. luku 5.2.), joten päätelmä ei täysin pidä paikkaansa. Sen sijaan kielitajun käyttämättömyydelle tällä tasolla voi esittää muutaman perustelun. Ensinnä teksteihin ja puhuttuihin kielen muotoihin keskittyvä tutkimus pitää itse tärkeänä, että aineisto on koostettu aidosta kielenkäytöstä. Näissä suuntauksissa kielitajun varaisen esimerkistön käyttö kohtaa luultavasti kritiikkiä. Toiseksi puhutut ja kirjoitetut tekstit ilmenevät ja niitä myös pidetään kielenkäyttötapauksina, joiden sepittäminen kielitajun varassa olisi looginen virhe: tutkija sepittäisi ensin teoreettista väittämäänsä vastaavan tekstin ja esittäisi sitten saman tekstin todellisena esimerkkitapauksena kielenkäytöstä. Tällaisen välttäminen ei ole kielitajun rajallisuutta metodina vaan pikemmin asiaankuuluvaa pidättäytymistä argumentointivirheestä.

Kuten minkä tahansa metodin myös kielitajun puutteita ja rajoitteita tulee avoimesti pohtia. Vähintään yhtä tärkeää olisi tutkia erilaisia metodeja itseään ja esimerkiksi kielitajun luonnetta vieläkin tarkemmin, jotta ymmärrys tutkimuksen luonteesta tarkentuisi. Tällä tavoin yksittäisten metodien käyttö ja keskinäinen yhdistely muodostuisivat yhä tarkoituksenmukaisemmaksi. On vielä todettava, että kielitaju on toimiva metodi omalla vahvuusalueellaan. Se on myös välttämätön osa mitä tahansa kielentutkimusta tutkimushypoteesien muodostelussa (esim. naturalistisen observoinnin pohjalta), monien metodien esi- tai osatöinä (esim. kysely, kokeelliset menetelmät, elisitointi) sekä tulosten tulkinnassa (ks. esim. Arppe 2008, 1–7). Kaikkein empiirisinkään tutkimus ei voi pidättyä kokonaan kielitajun käytöstä, koska esimerkiksi aineistoa ei voi kerätä ilman alustavaa käsitystä tutkittavasta ilmiöstä.

# 5. Kielitaju käytännössä

Tutkijan kielitaju on tärkeä työkalu ja erottamaton osa kaikenlaista kielentutkimusta. Erityisen tärkeä se on silloin, kun tarkastellaan yksittäisten ilmausten merkityksiä. Tässä luvussa keskitymme tähän ja tarkastelemme kielitajun käyttöä sellaisessa kielentutkimuksessa, joka käyttää tutkimusaineistona joko pelkästään tai osittain tutkijan itsensä sepittämiä esimerkkejä. Luvussa 5.1. esittelemme kielitajun avulla hankitun aineiston luonnetta. Luvussa 5.2. selostamme Talmyn (2007a) esitystä kielitajun kyvystä tavoittaa erilaisia kielen ilmiöitä. Luvussa 5.3. havainnollistamme esimerkkitapauksella kielitajun käyttöä aidossa tutkimuksessa.

## 5.1. Data, evidenssi, teoria?

Kielentutkimuksessa kielitaju on tietoista huomion kohdistamista erilaisiin kielenpiirteisiin sellaisina, kuin ne ilmenevät tutkijan kognitiossa (Talmy 2007b, xii). Karlsson (2008, 14) erittelee menetelmän käyttöä tarkemmin siten, että siihen kuuluu yhtäältä ensi vaiheessa äidinkielisen kielitajun konsultoiminen ylipäätään ja toisaalta täysin kieliopillisten (selvien tapausten), epäkieliopillisten ja ei-yksiselitteisten tai kyseenalaisten lause-esimerkkien tuottaminen. Kielitajun ei tietenkään tarvitse rajoittua vain lauseisiin. On myös huomattava, että sillä tuotettu kielenaines ei ole ”kompetenssiaineistoa”, sillä abstraktiin chomskylaiseen kompetenssiin ei ole suoraa pääsyä (Penke & Rosenbach 2007, 14). Kielitaju ilmentää myös kielellistä performanssia. Näin tuotettu kielenaines poikkeaa silti eräiltä osin merkittävästi tyypillisen empiirisen metodin keinoin hankitusta aineistosta. Herääkin kysymys, missä määrin kielitajulla saatu aines poikkeaa tyypillisestä empiirisestä aineistosta ja mitä tieteenfilosofisia seuraamuksia tällä mahdollisesti on.

Ensinnä on huomattava, että kielitajun tuottama kielenaines on refleksiivistä, ikään kuin itsestään tietoista, ts. se ei olisi olemassa ilman käyttötarkoitustaan tutkijan itsensä esittämänä havaintoaineistona jostain

kieltä koskevasta seikasta. Tässä suhteessa tällainen kielenaines eroaa jyrkästi spontaanista ja autenttisesta kielenkäytöstä, joka usein erilaisin menetelmällisin keinoin erikseen saatetaan tutkimusaineistoksi (tallentaminen, litterointi, kieliopillinen koodaus jne.). Tästä prosessista antaa valaisevan kuvan Lehmann (2004). Huomattakoon, että myös autenttinen kielenaines on refleksiivistä mutta eri tavalla: tilanteisessa kielenkäytössä kielellisten valintojen tarkoituksenmukaisuutta arvioidaan ja tulkitaan vuorovaikutuksellisista lähtökohdista, osana meneillään olevan toiminnan mielekkyyttä ja rakentumisehtoja. Tällöinkin kielioppia monitoroidaan mutta ei yleensä kielentutkimuksellisella motivaatiolla.

Toiseksi ulkoinen aineisto koostuu esiintymistä (engl. *token*), kun taas kielitajulla tuotettu aines edustaa tyyppisyyttä (engl. *type*). Sillä ilmennetään yleensä jonkin kieliopillisen rakenteen tilanteisesta käytöstä riippumatonta lingvististä ilmiötä yleistetyllä tyyppien tasolla. Kielitajun käyttö aineslähteenä onkin tyypillisesti kiinteä osa teoreettista kielenkuvausta. Konstenius (2014, 151) kuvaa kielitajun mukaisten esimerkkien luonnetta osuvasti: ”Teoreettisen kieliopin heuristiset esimerkit, joiden tarkoitus on osoittaa lukijalle kirjoittajan jokin tarkoite tai intuitio, toimittavat ostensiivista roolia. Johtopäätös ei perustu esimerkkiin. Esimerkkien tehtävä on vain osoittaa lukijalle kirjoittajan tarkoittamia rakenteita.” Sen sijaan tyypillinen korpusaineisto tai ulkoinen raakadata eli aineistoksi vielä muokkaamaton mutta tallennettava kielenaines ei sisällä tällaista episteemistä elementtiä. Kielitajun mukaiset esimerkit ovat tässä mielessä autonomisia silloinkin, kun niillä ilmennetään jotain kielenkäyttöön ja kontekstiin liittyvää tyyppisyyttä. (Ks. myös Devitt 2006.)

Kolmanneksi ja edellisen seurauksena kielitajulla ei voi tuottaa aineistoa, sitten tallentaa sitä ja tämän jälkeen tuoda sitä tutkittavaksi ikään kuin empiirisenä aineistona. Tällaiseen ei yleensä langetakaan, mutta kielitajulla tuotettujen tekstiesimerkkien osalta voi helposti herätä tällainen (joskus perusteeton) kritiikki. Kuitenkin voidaan tuottaa myös lausetta pidempiä (teksti)esimerkkejä esimerkiksi havainnollistettaessa erilaisia lauserajat ylittäviä ilmiöitä, kuten pronominien viittauskäytäntöjä (ks. esim. Larjavaara 2007, 482). Tätä tapahtuu piilevästi myös esimerkiksi silloin, kun irtolauseiden taustalle ajatellaan niiden potentiaalisia ja mahdollisia esiintymiskonteksteja. Tämä ei kuitenkaan

kielitajumetodin sisällä implikoi eikä siitä seuraa sitä, että tutkittaisiin todellista kielenkäyttöä. Kielitaju vain soveltuu joiltain osin myös tekstuaalisiin ilmiöihin, sillä moniin tekstitason tyyppi-ilmiöihin pääsy on kielitajun varassa. Tuotettujen esimerkkien poikkeavuus ulkoisesta aineistosta saa Lehmannin (2004, 292) pitämään kielitajua lähinnä tutkijan heuristisena työkaluna, joka jää empiirisen metodologian ulkopuolelle. Näin tuotetuilla esimerkeillä ei myöskään tällöin olisi aineiston asemaa tavanomaisten empiiristen metodien puitteissa.

## 5.2. Leonard Talmyn online- ja offline-introspektio

Kielitajua koskevien ilmiöiden yksityiskohtaisessa analyysissaan Talmy (2007a; 2007b) lähtee siitä, että (äidin)kieltä koskeva itsehavainnointi ei ole pelkästään lingvistiikan tutkimusmetodi vaan arkipäiväinen ja kaikille ihmisille yhteinen kyky ja taipumus, jota jokainen harjoittaa jatkuvasti monella eri tavalla.[15] Tätä kykyä kielentutkija käyttää erikseen harjaannutettuna hyväkseen.

Talmyn mukaan introspektio toimii tietyillä kielen alueilla paremmin ja toisilla huonommin, ja eräät kielenilmiöt ovat käytännössä sen ulottumattomissa. Talmy myös erottaa toisistaan kielenkäyttötilanteessa tapahtuvan introspektion (engl. *online introspection*) ja kielenkäytöstä irrallaan tapahtuvan introspektion (engl. *offline introspection*). Näiden tehtävät eroavat toisistaan, ja niiden avulla voidaan myös tarkastella erilaisia kielenilmiöitä. Kielentutkija, joka käyttää kielitajuaan analyysivälineenä, turvautuu yleensä ensisijaisesti offline-introspektioon.

Offline-introspektiolle suhteellisen helposti avautuvia seikkoja ovat Talmyn mukaan esimerkiksi tietyn sanan perusmerkitys (*kissa*, *laulaa*, *punainen*), ilmauksen edustama rekisteri (*minä menen* vs. *mä meen*) ja sen tyylillinen sopivuus tiettyyn kontekstiin, kieliopillisuus tai kielenvastaisuus (*sinä menet* vs. **sinä menen*), synonymia, samaan merkityskenttään kuuluminen (esim. työkalujen nimitykset) ja vastakkaismerkityksisyys (*matala–syvä*). Luettelosta voi huomata, että siihen kuuluu ennen kaikkea merkitykseen liittyviä kielenilmiöitä. Talmy (2007a) korostaakin, että juuri merkityksen tutkimus on kielitieteen ala,

jota ei voi harjoittaa käyttämättä introspektiota (eli kielitajua). Ilmiöinä, jotka avautuvat vain jonkin verran offline-introspektiolle, Talmy mainitsee kieliopillisten morfeemien merkitykset, sanaluokkajaon sekä ilmauksen muodon tai merkityksen salienssin (silmiinpistävyyden). Huonosti tai ei lainkaan offline-introspektion tavoitettavissa taas ovat muun muassa useiden kieliopillisten elementtien semanttiset komponentit (esim. liitepartikkelin -*kin* merkitys), polyseemisen ilmauksen kaikki merkitykset, syntaktiset säännöt ja muodot sekä vaikkapa luettelon esittäminen sanoista, joissa on tietty fonologinen piirre (esim. viidentenä äänteenä *r*).

Talmy (2007b) jaottelee vastaavasti tekijöitä, jotka avautuvat aidossa kielenkäyttötilanteessa tapahtuvan online-introspektion avulla joko helposti, jonkin verran tai eivät ollenkaan. Kielitajun online-käyttö on olennainen osa kielentutkijoiden usein harjoittamaa naturalistista observointia, jossa tehdään havaintoja aidosta kielenkäytöstä. Suhteellisen helposti avautuvia ilmiöitä ovat muun muassa keskustelun aihepiiri, keskustelun koherenssin taso, puhujan parhaillaan ilmaisema käsitesisältö, (puhujan kannalta) sanottavan sisällön vielä ilmaisematta oleva osa, puheenvuoroa suunnittelevan kannalta sisältö, joka aiotaan ilmaista, kuulijan kannalta puhujan poikkeamat tyylistä ja hyvämuotoisuudesta sekä virheet ja takeltelu. Jonkin verran avautuvia seikkoja ovat paralingvistiset ilmiöt eli sellaiset puheessa ilmenevät oheisilmiöt kuin puheen korkeus, voimakkuus, nopeus, sekä (puhujan kannalta) poikkeamat, jotka mahtuvat käytössä olevan rekisterin sisälle (ks. Kokeellinen fonetiikka tk.). Seikkoja, jotka eivät lainkaan avaudu online-introspektiolle, ovat muun muassa käynnissä olevat, puheen tuottamiseen tai ymmärtämiseen liittyvät kognitiiviset prosessit.

## 5.3. Lausesemantiikan esimerkkitapaus

Tutkijan itsensä konstruoimilla esimerkeillä operoiva tutkimus on vahvimmillaan silloin, kun tutkimuskysymys on muodoltaan sillä tavoin yleinen tai abstrakti, että sitä olisi vaikea palauttaa tietyn konkreettisen kielellisen muodon esiintymien tarkasteluun. Tällaiset pohjaavat usein teoreettisiin kysymyksiin. Käsittelemme seuraavaksi esimerkkinä

Huumon (2006) tarkastelemaa tutkimuskysymystä siitä, miten suomen objektin sijanvaihtelu toimii lauseissa, joissa joko subjektin tai objektin tarkoitteena on ns. inkrementaalinen teema (termistä ks. Dowty 1991). Inkrementaalinen teema on tarkoite, joka osallistuu tilanteeseen vaiheittain niin, että tilanteen edetessä ajassa yhä suurempi osa tarkoitteesta on joutunut verbin ilmaiseman toiminnan kohtaamaksi. Tilanteen ajallinen kesto korreloi suoraan inkrementaalisen teeman tarkoitekvantiteetin kanssa: kun koko kvantiteetti on osallistunut tilanteeseen, tilanne päättyy. Esimerkkinä on lause *Pekka söi omenaa ~ omenan*. Siinä objektin tarkoite on inkrementaalinen teema, sillä omena syödään osa kerrallaan ja tilanne kestää niin kauan kuin omenaa riittää. Kun koko omena on syöty, tilanne päättyy. Suomen objektin sijamerkinnällä tunnetusti ilmaistaan (eräiden muiden seikkojen ohessa) sitä, onko kuvattu tilanne kehityksenalainen vai päättynyt: partitiiviobjekti (*omenaa*) esittää tilanteen kehityksenalaisena, akkusatiiviobjekti (*omenan*) päättyneenä. Silloin kun objekti on inkrementaalinen teema, akkusatiiviobjekti kertoo samalla, että objektin koko tarkoite on osallistunut tilanteeseen (esim. koko omena on tullut syödyksi).

Objektin sijamerkinnän osalta inkrementaalisen teeman käyttäytyminen on ollut fennistiikassa tunnettu ilmiö jo pitkään. Huumo (2006) ryhtyykin pohtimaan sitä, miten objekti ja subjekti merkitään silloin, kun inkrementaalinen teema esiintyy subjektina. Tällöin tilanteen käynnissä oleminen tai päättyminen riippuu siitä, onko subjektin koko tarkoite jo ehtinyt osallistua siihen vai ei. Merkitäänkö myös subjektitarkoitteen osallistumisen vaiheittaisuus kieliopillisesti jollain tavoin, ja jos niin miten?

Näin muotoiltuna kysymyksen ratkaiseminen korpuksen avulla olisi hankalaa. Inkrementaalinen teema on abstrakti semanttinen rooli, joka ei korreloi suoraan minkään leksikaalisen tai kieliopillisen muotopiirteen kanssa niin, että sitä voisi esimerkiksi automaattisesti etsiä korpuksesta tavanomaisin hakukriteerein. On tietenkin mahdollista käydä manuaalisesti läpi tekstiaineistoa ja yrittää löytää transitiivilauseita, joiden subjektin tarkoite olisi inkrementaalinen teema. Helpoin ja kätevin ratkaisu on kuitenkin käyttää tutkijan omaa kielitajua, yrittää sepittää tällaisen ehdon täyttävä, kielessä hyväksyttävä lause ja tarkastella sitä.

Huumo (2006) esittää tämäntyyppisen esimerkkilauseen: *Tavarajuna sivuutti opastinta ~ opastimen*. Esimerkissä objektin tarkoite (opastin) on pistemäinen ja kokonaisuutena mukana tilanteessa sen joka vaiheessa. Sitä vastoin subjektin tarkoite (tavarajuna) on pitkänomainen ja osallistuu opastimen sivuuttamiseen osa kerrallaan niin, että opastimen kohdalla sitä sivuuttamassa on kullakin hetkellä vain jokin junan osa (esimerkiksi vaunu). Esimerkin objektin sija ja mahdollisuus valita joko partitiivi- tai akkusatiiviobjekti osoittavat, että merkintäperiaate on silti sama kuin aiemmassa esimerkissä, jossa inkrementaalinen teema oli objektina ('syödä omenaa'). Jos tilanne esitetään kehityksenalaisena tai keskeneräisenä, käytetään partitiiviobjektia (*opastinta*), jos taas päättyvänä, käytetään akkusatiiviobjektia (*opastimen*). Esimerkki osoittaa, että subjektitarkoitteenkin inkrementaalisuudesta seuraava tilanteen kehityksenalaisuuden ja päättymisen oppositio ilmaistaan objektin sijalla, vaikka objekti itse ei nyt olekaan inkrementaalinen teema.

Kielitajua käyttävä kielentutkija voi jatkaa pohdintaa kysymällä, pystyisikö subjektin (*tavarajuna*) oma sijanvalinta heijastelemaan tällaista eroa: voisihan ajatella, että kun juuri subjektin tarkoite on inkrementaalinen teema, sijanvaihtelu merkittäisiinkin subjektiin eikä objektiin. Vastaus on kuitenkin kielteinen, sillä tällaista merkintää tavoitteleva partitiivi lauseessa **Tavarajunaa sivuutti opastimen* on selvästi kielenvastainen.[16] Asian osoittaminen on oikeastaan mahdollista vain kielitajun ja itse tuotetun sekä kielenvastaiseksi havaitun esimerkin avulla, sillä negatiivista evidenssiä eli tällaisen lauseen esiintymättömyyttä ei voi korpuksista helposti havainnoida. Se, että tämäntyyppistä lausetta ei (todennäköisesti) löytyisi mistään aineistosta, ei vielä edes riittäisi osoittamaan ilmauksen kielenvastaisuutta.

Subjektin ja objektin sijan tarkastelu tavarajuna-esimerkissä havainnollistaa nominatiivin ja partitiivin erilaista **distribuutiota** eli jakaumaa tällaisessa ilmaustyypissä. Kielenilmiön ja jonkin kielellisen aineiston analysointia esimerkiksi minimipareilla ja rakenteiden käyttöalaa koskevaa distribuutiota tarkastelemalla voi pitää konkreettisena metodisena toimintana. Aineistonhankintametodeista (esim. kysely, haastattelu jne.) tällainen eroaa siten, että kyse on jo usein teoreettisesta argumentoinnista ja jonkin ilmiön kuvaamisesta.

Palataksemme esimerkkiin tarkoitettua asiasisältöä kuvaisi kyllä esimerkiksi lause *Osa tavarajunaa sivuutti opastimen*, mutta tämä ei edustaisi tavoiterakennetta. *Tavarajuna* on partitiivissa sanan *osa* jälkimääritteenä (*osa leipää, tarinaa* yms.). Minimipariesimerkeillä operoimisessa on tärkeää pitäytyä saman rakennetyypin piirissä. Myös esimerkki *Tavarajuna ohitti matkustajajunaa ~ matkustajajunan* on siksi toimimaton, että se muuttaisi alkuperäistä asiaintilaa, jossa ohitettava on pistemäinen tarkoite (opastin). Jos kielitajun varaisesta lause-esimerkistä tulee kiistaa, aina on käytännöllistä yrittää etsiä korpusesimerkkejä: *Puomi oli ollut alhaalla ja ääni- ja valomerkit toiminnassa koko ajan, kun tavarajuna ohitti risteystä ja kun pikajuna oli saapunut hälytysosuudelle.* [17]

Kielitajuaan hyödyntävän kielentutkijan ei kannata tyytyä vain yhteen esimerkkiin, sillä aina on mahdollista, että havaitun eron takana on muitakin tekijöitä kuin se, joka oli huomion kohteena. On pohdittava monenlaisia ehdot täyttäviä lauseita ja yritettävä aktiivisesti keksiä omalle hypoteesille ja sitä tukevalle esimerkille vastaesimerkkejä. Kieliopillisten seikkojen lisäksi esimerkiksi pragmaattiset tai leksikaaliseen semantiikkaan liittyvät seikat saattavat vaikuttaa lauseiden hyväksyttävyyteen, joten näitäkin tekijöitä on hyvä yrittää muunnella.

Aina on hyvä myös esittää jatkokysymyksiä ja lisähypoteeseja: entäpä jos sivuuttamisen tyyppiseen tilanteeseen vaiheittain osallistuva pitkänomainen subjektitarkoite onkin kvantiteetiltaan rajaamaton eikä rajattu niin kuin tavarajuna? Huumo (2006) esittää esimerkin kerjäläisestä, jonka ohi vaeltaa turistien loputon (ja siten kvantitatiivisesti rajaamaton) virta. Tällaista tilannetta transitiivisella *sivuuttaa*-verbillä kuvattaessa huomataan, että akkusatiiviobjekti onkin mahdollinen riippumatta siitä, onko tilanne kehityksenalainen vai ei. Lause *Turistien loputon virta sivuutti kerjäläisen* sallii akkusatiiviobjektin siinäkin tapauksessa, että se kuvaa käynnissä olevaa sivuuttamista. Toisaalta partitiivi tuntuu esimerkissä *?Turistien loputon virta sivuutti kerjäläistä* vastahankaiselta juuri siksi, että tilanne ei etene kohti päätepistettä, jolloin se päättyisi, sillä turistivirta esitetään loputtoman pitkänä.[18] Samantapaisia ovat lauseet *Joessa virtaava vesi kiertää nykyään kaupungin keskustan kanavaa pitkin* ja *Liikennevirta ohitti onnettomuuspaikan kiertotietä pitkin*. Molemmissa subjektitarkoite on ainesmäärältään rajaamaton. Voidaankin havaita,

että objektin sijanvalinta toimii hieman eri tavoin silloin, kun tilanteeseen vaiheittain osallistuva subjektitarkoite on ainesmäärältään rajallinen kokonaisuus (tavarajuna), ja silloin, kun tämä on ainesmäärältään rajaamaton (turistien virta, vesi). Tällaiseen havaintoon voi päätyä kyseenalaistamalla tavarajuna-esimerkin ja koettamalla keksiä sille vastaesimerkkejä. Tavarajuna-esimerkin perusteella tehty päätelmä siis tarkentuu rajaamaton ainesvirta -tyyppisten esimerkkien avulla.

## 6. Yhteenveto

Kielitajun käyttö on olennainen osa lähes kaikkea kielitieteellistä tutkimusta. Silloinkin kun varsinainen tutkimusaineisto on aidosta kielenkäytöstä peräisin, tutkija yleensä turvautuu kielitajuunsa ainakin tutkimuskysymyksiä ja hypoteeseja muotoillessaan, aineistoa luokitellessaan, tutkimustuloksia tulkitessaan ja päätelmiä tehdessään. Kielitaju on myös opinnäytettä kirjoittavalle opiskelijalle tärkeä, ellei tärkein työkalu. Kielitaju ei metodina ole ”epätieteellinen” tai ”mielivaltainen”, sillä kielellisen normin yhteisyys puhujayhteisön jäsenien kesken toimii takeena siitä, ettei mielivaltaisia väitteitä voi esittää. Kuten mitä tahansa metodia myös kielitajua on mahdollista käyttää hyvin tai huonosti: esimerkiksi Dąbrowskan (1996) kritisoimat tutkijat ovat kielitajunsa varassa muodostaneet hyvin kompleksisia esimerkkejä, jollaisia ei esiinny aidossa kielenkäytössä ja joita kaikki äidinkielisetkään puhujat eivät pysty ymmärtämään.

Hiukan toinen kysymys on se, milloin tutkimuksen aineistona voi ja kannattaa käyttää itse sepitettyjä esimerkkejä. Luvussa 5.3. kuvasimme pelkistetysti erään tällaisen tutkimuksen kulkua. Tutkimuskysymykset voivat tällaisessa tutkimuksessa olla muotoilultaan varsin abstrakteja. Vastauksen saamiseksi riittää periaatteessa tietynlaisen esimerkin tuottaminen ja sen hyväksyttävyyden tai kielenvastaisuuden toteaminen sekä yleensä myös sen merkitykseen vaikuttavien tekijöiden systemaattinen teoreettinen analysointi. Kielitajun varassa tuotettu esimerkki voi toimia

kirvokkeena tutkimuskysymysten täsmentämiseksi niin, että asiaa voidaan edelleen tutkia todellista kielenkäyttöä edustavasta aineistosta tai muilla kielentutkimuksen menetelmillä.

Itse tuotettujen esimerkkien keksiminen ja puntarointi, niiden merkitykseen vaikuttavien seikkojen erittely sekä mahdollisten vastaesimerkkien sepittäminen vaativat tuekseen harjaantumista ja analyyttista taitoa, jollainen yleensä kehittyy vasta melko pitkän tutkimuskokemuksen myötä (vrt. Talmyn vertaus urheilijaan). Ainakaan ensimmäisiä opinnäytteitään, kandidaatintutkielmaa tai gradua, laativalle opiskelijalle sitä ei voi suositella ensisijaiseksi tai yksinomaiseksi menetelmäksi. Kielitajun käyttöä tutkimusmenetelmänä on paras harjoitella ensin tosiasiallista kielenkäyttöä edustavalla aineistolla, joka opettaa ymmärtämään kielen ja sen käyttöön liittyvien tekijöiden moninaisuutta. Todellista kielenkäyttöä edustavaa aineistoa tutkivakin voi tietysti pohtia – ja usein pohtiikin – mitä tapahtuisi, jos ilmauksen yksi elementti korvattaisiin toisella, ja päätyä näin tuottamaan ”laboratoriolauseen”, joka toimii kielitajun varaisena minimiparina aidosta kielenkäytöstä poimitulle esimerkille.

**Aiheesta lisää:**


Devitt, Michael. 2006. Intuitions in linguistics. *The British Journal for the Philosophy of Science* 57, 481–513.

Labov, William. 1996. When intuitions fail. Julkaisussa: McNair, Lisa, Singer, Kora, Dolbrin, Lise & Aucon, Michelle (toim.) *Papers from the Parasession on Theory and Data in Linguistics*. Chicago Linguistic Society 32. Chicago (IL): Chicago Linguistic Society, 77–106.

Itkonen, Esa. 1981. The concept of linguistic intuition. Julkaisussa: Coulmas, Florian (toim.) *A Festschrift for Native Speaker*. The Hague: Mouton, 127–140.

Willems, Klaas. 2012. Intuition, introspection and observation in linguistic inquiry. *Language Sciences* 34, 665–681.

## VIITTEET

1 Asteriski (*) on konventionaalinen tapa merkitä ei-hyväksyttävyyttä tai epäkieliopillisuutta.

2 Runous ja kielellinen leikittely ja jopa kielivirheet ovat toki ymmärrettävissä, mutta tuolloin taustalla on aina jokin normatiivinen tai muu kehys, jota vasten tulkintaa verrataan. Kielellinen leikittely ei yleensä muuta normijärjestelmää.

3 On toistaiseksi vielä erillisen teoreettisen tutkimuksen aihe pohtia sitä, rajaako tyypillinen kielitajun varainen yleiskielinen aineisto kielestä esitetyt väitteet koskemaan vain standardikielimuotoa vai voiko ja millä edellytyksillä kielitajun varassa sanoa muutakin kielestä. Yleisesti kielitieteellisillä aineistoilla voidaan perustella kielen universaaleja (kaikille tai monille kielille yhteisiä ominaisuuksia), jonkin yksittäiskielen ominaisuuksia, yksittäiskielen genren, rekisterin ja variaation piirteitä tai yksittäisen aidon kielenkäyttötilanteen ominaisuuksia. Kielitajulla ei havainnollisteta jälkimmäisintä.

4 Ensikielen tai äidinkielen rakennetta koskeva taito saavutetaan asteittain, joten kielitaidon mittana pidetään yleensä oletettua aikuiskielen kehittynyttä kielitajua. Tämä saavutetaan toisella kymmenellä ikävuodella.

5 Leikkikieli kontinkielestä on kuitenkin todettava, että se perustuu luonnolliselle kielelle (suomelle) tietyin muodostussäännöin, mutta kukaan ei opi sitä äidinkielenään. Leikkikielet ovat yleisesti harvinaistuneet.

6 Tarkoitamme ei-kielentutkijoilla tai (tavallisilla) kielenpuhujilla kaikkia niitä äidinkielisiä puhujia, joilla ei ole kielentutkijan koulutusta eikä ensisijaisesti tieteellis-teoreettista motivaatiota kielen tarkkailuun.

7 Omiin tajunnansisältöihin kohdistuvan itsehavainnoinnin lisäksi kielitajua lähellä oleva arkinen metodi kielentutkijalla on luonnollisen kielenkäytön ulkopuolinen tarkkailu tai eri astein osallistuva havainnointi eli nk. **naturalistinen observointi** (ks. online-introspektio, luku 5.2.). Tällainen voi satunnaisena toimintana synnyttää relevantteja tutkimushypoteeseja tai systemaattisena toimintana johtaa merkittävään aineistokeräelmään jostakin kielenkäytön ilmiöstä. (Ks. Karlsson 1998, 285–286.)

8 Muita kuin kieltä koskevia varmoja asioita kuvaavat esimerkiksi lauseet *olen elossa*, *minulla on sydän* ja *en ole käynyt koskaan Kuussa*. Tosiasiassa varmuus koskee Wittgensteinin mukaan tämän tyyppisten tiedonsisältöjen muodostamaa järjestelmää (§ 410).

9 Kieliopin säännön ”alaraja” lienee vedettävissä jonnekin fonologian ja fonetiikan rajamaastoon. Esimerkiksi suomen rajageminaatio (esim. *hernekeitto* [hernek:eitto]) voi olla monelle tietoisuuden alinen ilmiö, vaikka se toteutuu säännönmukaisesti monissa suomen muotoryhmissä. Toinen esimerkki voisi koskea esim. diftongin avartumisasteita (*hieno* ~ *hiano*). On kuitenkin huomattava, että tällaisten ilmiöiden tiedostamista ei ole kovin perusteellisesti kartoitettu.

10 Propositionaaliset tietosisällöt ovat väitelauseen muodossa esitettävissä olevia uskomuksia, esim. *olen tyytyväinen* (vrt. tyytyväisyyden kielentämätön mutta introspektiolla tavoitettava tunne kokijan mielessä) (Niiniluoto 1997 [1980], 137).

11 Esimerkissä ongelmana on Apollonioksen aikansa normaalikreikkaa koskevan kielitajun mukaan se, että modaalipartikkeli (*an*) ei esiintyisi indikatiivin futuurin (*grapso*) kanssa. Kuitenkin homeerisessa kreikassa partikkelin varianttia *ken* käytetään näin. (Minna Seppänen h.t. 29.1.2018.)

12 Vastanäkemyksenä introspektionismille psykologiassa kehittyi 1900-luvun alkupuolella behaviorismi lähestymistavaksi, jossa pätevää tietoa ajateltiin saatavan vain ulkoista käyttäytymistä havainnoimalla (ks. Lyons 1986, luku 2; positivismista ks. Johdannon luku 2 tk.).
13 Tesnière käyttää kielitajusta termiä *introspektio* (suom. toim.).
14 Suomessa *i* : *e* -nominit ovat niitä nomineja, joiden nominatiivi päättyy *i*:hin (*korpi*) ja taivutuksen vartalovokaali on *e* (*korve-*). Tähän taivutustyyppiin kuuluvat muun muassa sanat *sormi*, *ovi*, *savi* ja *kivi*.
15 Talmy käyttää kielitajusta yksinomaan termiä *introspektio*, jota tässä seurataan.
16 Esimerkiksi lauseen ***Monia tavarajunia*** *sivuutti opastinta ~ opastimen* partitiivisubjekti on käypä, sillä monikollinen kvanttorisana *monia* epämääräistää subjektitarkoitteiden määrän.
17 Yle Uutiset, 26.7.2001. https://yle.fi/uutiset/3-5107078.
18 Kysymysmerkki esimerkin edessä ilmentää esimerkin kyseenalaista suomenmukaisuutta (ks. kuvio 1, mahdollisesti p).

## KIRJALLISUUS


Apollonius Dyskolos. 1981. *The Syntax of Apollonius Dyscolus*. [*Peri syntakseōs*]. Kääntänyt ja kommentein varustanut Fred Householder. Amsterdam Studies in the Theory and History of Linguistic Science III. Studies in the History of Linguistics, Volume 23. Amsterdam: John Benjamins.

Arppe, Antti. 2008. *Univariate, Bivariate, and Multivariate Methods in Corpus-based Lexicography – a Study of Synonymy*. Väitöskirja. Helsinki: University of Helsinki.

Bloomfield, Leonard. 1983 [1914]. *An Introduction to the Study of Language*. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Bloomfield, Leonard. 1957 [1933]. *Language*. London: George Allen & Unwin Ltd.

Carroll, John B. 1955. *The Study of Language. A Survey of Linguistics and Related Disciplines in America*. (2. p.). Cambridge: Harvard University Press.

Chomsky, Noam. 2001 [1965]. *Aspects of the Theory of Syntax*. (21. p.). Cambridge (MA): MIT Press.

Chudnoff, Elijah. 2013. *Intuition*. Oxford: Oxford University Press.

Dąbrowska, Ewa. 1996. The LAD goes to school. A cautionary tale for nativists. *Linguistics* 35, 735–766.

Dąbrowska, Ewa. 2010. Naive v. expert intuitions: An empirical study of acceptability judgments. *The Linguistic Review* 27, 1–23.

Devitt, Michael. 2006. Intuitions in linguistics. *The British Journal for the Philosophy of Science* 57, 481–513.

Devitt, Michael. 2014. Linguistic intuitions and cognitive penetrability. Julkaisussa: Machery, Edouard & Prinz, Jesse (toim.) *The Baltic International Yearbook of Cognition, Logic and Communication* (Vol. 9), *Perception and Concepts*. [verkkoaineisto]. [viitattu 7.2.2019]. Saatavissa: https://newprairiepress.org/biyclc/vol9/iss1/.

Dowty, David. 1991. Thematic proto-roles and argument selection. *Language* 67:3, 547–619.

Droste, Flip G. & Joseph, John E. (toim.). 1991. *Linguistic Theory and Grammatical Description*. Current Issues in Linguistic Theory, Volume 75. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Fillmore, Charles J. 1992. ”Corpus linguistics” or ”Computer-aided armchair linguistics”. Julkaisussa: Svartvik, Jan (toim.) *Directions in Corpus Linguistics, Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991*. Trends in Linguistics. Studies and Monographs 65. Berlin: Mouton de Gruyter, 35–60.

Firth, John Rupert. 1957. *Papers in Linguistics 1934–1951*. London: Oxford University Press.

Fries, Charles C. 1970. The Bloomfield ’school’. Julkaisussa: Mormann, Christine, Sommerfelt, Alf & Whatmough, Joshua (toim.) *Trends in European and American Linguistics 1930–1960*. Utrecht: Spectrum Publishers, 196–224.

Gardiner, Alan. 1951 [1932]. *The Theory of Speech and Language*. (2. p.). Oxford: Clarendon Press.

Hakulinen, Auli & Karlsson, Fred. 1995 [1979]. *Nykysuomen lauseoppia*. (3. muuttamaton p.). SKST 350. Helsinki: Suomalaisen Kirjallisuuden Seura.

Hamunen, Markus, Nieminen, Tommi, Kelomäki, Tapani & Dufva, Hannele (toim.). 2019. *Käänteitä ja käsitteitä. Näkökulmia kielitieteelliseen keskusteluun*. SKST 1454. Helsinki: Suomalaisen Kirjallisuuden Seura.

Harris, Roy. 2009. *Rationality and the Literate Mind*. New York (NY) & London: Routledge.

Harris, Zellig S. 1969 [1951]. *Structural Linguistics*. Chicago (IL): The University of Chicago Press.

Helbig, Gerhard. 1973. *Geschichte de neueren Sprachwissenschaft. Unter dem besonderen Aspekt der Grammatik-Theorie*. Leipzig: Veb Bibliographisches Institut.

Henrich, Joseph, Heine, Steven J. & Norenzayan, Ara. 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33:2–3, 61–83.

Householder, Fred W. 1981. [Esipuhe]. Julkaisussa: *The Syntax of Apollonius Dyscolus* [*Peri Suntakseōs*]. Kääntänyt ja kommentein varustanut Fred Householder. Amsterdam Studies in the Theory and History of Linguistic Science III. Studies in the History of Linguistics, Volume 23. Amsterdam: John Benjamins, 1–17.

Huumo, Tuomas. 2006. Kvantiteetti ja aika I: nominaalisen aspektin näkökulma suomen objektin ja subjektin sijamerkintään. *Virittäjä* 110:4, 504–538.

Huumo, Tuomas. 2007. Joko lingvistin nojatuoli joutaisi kaatopaikalle? Introspektiolingvistiikan asemasta kognitiivisessa kielitieteessä. *Emakeele Seltsi aastaraamat* 53, 163–180.

Itkonen, Esa. 1974. *Linguistics and Metascience*. Studia Philosophica Turkuensia Fasc II. Turku: Turun yliopisto.

Itkonen, Esa. 1978. *Grammatical Theory and Metascience. A Critical Investigation into the Methodological and Philosophical Foundations of ’Autonomous’ Linguistics*. Amsterdam Studies in the Theory and History of Linguistic Science. Volume 5. Amsterdam: John Benjamins B. V.

Itkonen, Esa. 2003. *What is Language? A Study in the Philosophy of Linguistics*. Publications in General Linguistics 8. Turku: University of Turku.

Itkonen, Esa & Pajunen, Anneli. 2010. *Empiirisen kielitieteen metodologia. Suomi* 199. Helsinki: Suomalaisen Kirjallisuuden Seura.

Jantunen, Jarmo Harri. 2009. Ei pelkästään mielikuvituksen puutteen vuoksi: Kieliaineistojen systemaattinen käyttö kielentutkimuksessa. *Virittäjä* 113:1, 101–113.
Kantor, J. R. 1952 [1936]. *An Objective Psychology of Grammar*. Bloomington (Ind.): The Principia Press Inc.
Karlsson, Fred. 1982a. *Suomen kielen äänne- ja muotorakenne*. Porvoo: WSOY.
Karlsson, Fred. 1982b. *Suomen peruskielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Karlsson, Fred. 1998. *Yleinen kielitiede*. Uudistettu laitos. Helsinki: Yliopistopaino.
Karlsson, Fred. 2008. Early generative linguistics and empirical methodology. Julkaisussa: Kytö, Merja & Lüdeling, Anke (toim.) *Handbook on Corpus Linguistics*. Vol. 1. Berlin & New York (NY): Mouton de Gruyter, 14–32.
Kelomäki, Tapani. 2014. E. N. Setälän käsityksiä nuorgrammaattisen kielitieteen ja kieliopin kirjoittamisen suhteesta. *Virittäjä* 118:2, 263–272.
Kess, Joseph F. 1983. Introduction. Julkaisussa: Bloomfield, Leonard *An Introduction to the Study of Language* [1914]. Amsterdam & Philadelphia: John Benjamins Publishing Company, xvii–xxxviii.
Koksvik, Ole. 2011. *Intuition*. Väitöskirja. [verkkoaineisto]. [viitattu 6.2.2019]. Australian National University. Saatavissa: https://philpapers.org/archive/KOKI.pdf.
Koksvik, Ole. 2013. Intuition and conscious reasoning. *The Philosophical Quarterly* 63:253, 709–715.
Koksvik, Ole. 2017. The phenomenology of intuition. *Philosophy Compass* 12:1. [verkkoaineisto]. [viitattu 6.2.2019]. Saatavissa: https://onlinelibrary.wiley.com/doi/full/10.1111/phc3.12387.
Konstenius, Reetta Alexandra. 2014. *Empiria, eksperimentti ja etiikka. Kielitieteen metateoriaa*. Väitöskirja. Helsinki: Helsingin yliopisto.
Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Volume I. Theoretical Prerequisites*. Stanford (CA): Stanford University Press.
Larjavaara, Matti. 2007. *Pragmasemantiikka*. SKST 1077. Helsinki: Suomalaisen Kirjallisuuden Seura.
Lehmann, Christian. 2004. Data in linguistics. *The Linguistic Review* 21, 275–310.
Linell, Per. 2005. *The Written Language Bias in Linguistics. Its Nature, Origins and Transformations*. London: Routledge.
Ludlow, Peter. 2009. Having linguistic rules and knowing linguistic facts. *Baltic International Yearbook of Cognition, Logic and Communication* 5. Saatavissa: https://doi.org/10.4148/biyclc.v5i0.285.
Lyons, William. 1986. *The Disappearance of Introspection*. Cambridge (MA): MIT Press.
Maynes, Jeffrey & Gross, Steven. 2013. Linguistic intuitions. *Philosophy Compass* 8:8, 714–730.
Mills, Andrew P. [ei vuotta]. Knowledge of language. *The Internet Encyclopedia of Philosophy*. [verkkoaineisto]. [viitattu 7.2.2019]. Saatavissa: https://www.iep.utm.edu/knowlang/.
Moravcsik, Edith A. & Wirth, Jessica R. (toim.). 1980. *Current Approaches to Syntax*. Syntax and Semantics. Volume 13. New York (NY): Academic Press.
Mäkilähde, Aleksi, Leppänen, Ville & Itkonen, Esa (toim.). 2019. *Normativity in Language and Linguistics*. Studies in Language Companion Series 209. Amsterdam & Philadelphia: John Benjamins Publishing Company.

Niiniluoto, Ilkka. 1997 [1980]. *Johdatus tieteenfilosofiaan. Käsitteen- ja teorianmuodostus*. Helsinki: Otava.

Ojutkangas, Krista. 2008. Kielentutkijan mielikuvituksen puute eli mihin todelliseen kielenkäyttöön perustuvaa aineistoa tarvitaan. *Virittäjä* 112:2, 276–284.

Penke, Martina & Rosenbach, Anette. 2007. What counts as evidence in linguistics? An introduction. Julkaisussa: Penke, Martina & Rosenbach, Anette (toim.) *What Counts as Evidence in Linguistics. The Case of Innateness*. Benjamins Current Topics, volume 7. Amsterdam & Philadelphia: John Benjamins Publishing Company, 1–49.

Piippo, Irina, Vaattovaara, Johanna & Voutilainen, Eero. 2016. *Kielen taju: vuorovaikutus, asenteet ja ideologiat*. Helsinki: Art House.

Preston, Dennis R. 1996. Whaddayaknow. The modes of folk linguistic awareness. *Language awareness* 5:1, 40–73.

Raami, Asta. 2015. *Intuition Unleashed. On the Application and Development of Intuition in the Creative Process*. Väitöskirja. Media Lab, Department of Media, Aalto University, School of Arts, Design and Architecture.

Robins, R. H. 1969 [1967]. *A Short History of Linguistics*. London: Longmans, Green and Co Ltd.

Rorty, Richard. 1967. Intuition. Julkaisussa: Edwards, Paul (toim.) *The Encyclopedia of Philosophy*. Volume IV. New York (NY): The Macmillan Company & The Free Press, 204–212.

Saussure, Ferdinand de. 2014 [1916]. *Yleisen kielitieteen kurssi* [*Cours de linguistique générale*]. Toimittaneet Charles Bally ja Albert Sechehaye yhteistyössä Albert Riedlingerin kanssa. Suomentanut Tommi Nuopponen. Esipuhe Carita Klippi. Tampere: Vastapaino.

Schütze, Carson T. 1996. *The Empirical Base of Linguistics. Grammaticality Judgments and Linguistic Methodology*. Chicago (IL): The University of Chicago Press.

Spencer, N. J. 1973. Differences between linguists and nonlinguists in intuitions of grammaticality-acceptability. *Journal of Psycholinguistic Research* 2:2, 83–98.

Talmy, Leonard. 2003 [2000]. *Toward a Cognitive Semantics. Volume 1: Concept Structuring Systems*. Cambridge (MA): MIT Press.

Talmy, Leonard. 2007a. Introspection as a methodology in cognitive linguistics. Plenum-esitelmä ja ojenne, The 10th International Cognitive Linguistics Conference, Krakova, Puola 15.–20.7.2007.

Talmy, Leonard. 2007b. Foreword. Julkaisussa: Gonzalez-Marquez, Monica, Mittelberg, Irene, Coulson, Seana & Spivey, Michael J. (toim.) *Methods in Cognitive Linguistics*. Human Cognitive Processing 18. Amsterdam: John Benjamins, xi–xxi.

Tesnière, Lucien. 1959. Éléments de syntaxe structurale. Paris: Librairie C. Klincksieck.

Tieteen termipankki. 2019. [verkkoaineisto]. [viitattu 1.3.2019]. Saatavissa: https://tieteentermipankki.fi.

Timofeeva, Maria K. 2006. Introspective and traditional views of language. *Logic and Logical Philosophy* 15, 217–237.

Wasow, Thomas & Arnold, Jennifer. 2005. Intuitions in linguistic argumentation. *Lingua* 115, 1481–1496.

Willems, Klaas. 2012. Intuition, introspection and observation in linguistic inquiry. *Language Sciences* 34, 665–681.

Wittgenstein, Ludwig. 1999 [1969]. *Varmuudesta*. [Űber Gewissheit]. (2. p.). Suomentanut Heikki Nyman. Porvoo: WSOY.

# Luku 2 Kysely

*Ulla Vanhatalo*

https://orcid.org/0000-0002-2939-7836

*Kimmo Vehkalahti*

https://orcid.org/0000-0001-6540-7328

**Mikä?**

Kysely on keino tutkia, mitä ihmiset ajattelevat, tietävät, tekevät tai kokevat. Tiedon keruu ja mittaus tapahtuvat tutkimusta varten tehdyllä kyselylomakkeella, jonka laatiminen edellyttää sekä kielentutkimuksen että tilastotieteen osaamista. Mittausten tuloksena syntyy aineistoja, joita analysoidaan pääosin tilastollisilla menetelmillä. Kyselyllä kielestä voidaan saada sellaista uutta tietoa, jota on vaikeaa tai mahdotonta tavoittaa muilla keinoilla.

**Katso myös:**

Johdanto

Luku 1 Kielitaju kielentutkijan työkaluna

Luku 3 Haastattelu aineistonkeruun metodina sosiolingvistiikassa

Luku 16 Antropologinen lingvistiikka

Luku 23 Käännöstiede ja sen menetelmät

# 1. Johdanto

Kyselyn käyttöön opastetaan lukuisissa yhteiskunta- ja käyttäytymistieteiden tutkimusoppaissa (esim. Alkula, Pöntinen & Ylöstalo 1994; Foddy 2001; Fowler 1995; Saris & Gallhofer 2007; Vastamäki 2015; Vehkalahti 2019). Tässä artikkelissa tarkastelemme kyselyä kielentutkimuksen menetelmänä. Esittelemme kyselyn pääperiaatteita ja annamme käytännön ohjeita toimivan kyselyn tekemiseen. Aiheesta syvällisemmin kiinnostunutta ohjaamme menetelmän tärkeimpien lähteiden äärelle kielitieteen ulkopuolelle.

Kehotamme lukijaa pohtimaan erityisesti kolmea tutkimuksen onnistumisen kannalta olennaista kohtaa: 1) Miksi kysyt? Mikä on se tutkimuskysymys, johon haluat kyselyllä saada vastauksen? 2) Mitä kysyt? Mitä sinun on kysyttävä, jotta pystyt vastaamaan tutkimuskysymykseesi? 3) Miten kysyt? Millaisia tilastollisia ja muita näkökohtia sinun on huomioitava, jotta tutkimustuloksesi olisivat mahdollisimman luotettavia?

Kysely on keino tutkia mittaamalla, mitä ihmiset ajattelevat, tietävät, tekevät tai kokevat. Mittaus tarkoittaa sitä, että tutkittavasta ilmiöstä kerätään aineistoa, joka koostuu pääasiassa numeroista. Määrällinen tutkimus, mukaan lukien kysely, edellyttää useimmiten juuri numeerista mittausta. Näin kerättyjä aineistoja analysoidaan erilaisilla määrällisillä eli tilastollisilla menetelmillä. Kyselyllä voidaan kerätä myös sanallisia vastauksia, mutta sanalliset ja muut laadulliset aineistot on usein parempi analysoida laadullisilla menetelmillä. Tässä artikkelissa tarkastelemme kyselyä sen tyypillisessä käytössä numeromuotoisen kieliaineiston keruussa ja analysoinnissa, joskin monet näkökulmat sopivat myös laadulliseen kyselyaineistoon.

Erilaisia ilmiöitä mitataan erilaisilla mittareilla: vaa'alla mitataan painoa, mittanauhalla pituutta, sykemittarilla sydämen sykkeen taajuutta. Paino, pituus ja syke ovat suhteellisen yksinkertaisia mitattavia, ja mainitut mittarit tuottavat selviä numeroarvoja. Ihmisten tiedot, käsitykset ja asenteet ovat sen sijaan varsin monimutkaisia ilmiöitä, eikä niiden mittaukseen ole valmiita laitteita. Mittaus voi kuitenkin onnistua kyselylomakkeella. Kyselylomaketta voikin ajatella kokoelmana mittareita,

joilla mitataan ihmisen käsityksiä mitä erilaisimmista asioista. Ihmisen ajatteluun liittyvissä tutkimuksissa mittarit on usein rakennettava jokaista tutkimuskysymystä varten uudelleen, joskin toisinaan voidaan hyödyntää valmiita, aiemmin testattuja mittareita. Kyselylle on tyypillistä, että kaikki tutkimukseen valitut ihmiset vastaavat samoihin kysymyksiin. Kiinnostuksen kohteena eivät niinkään ole yksittäisen ihmisen vastaukset vaan kaikkien vastausten perusteella tehtävät tiivistykset, jotka heijastavat tutkittavaa ilmiötä.

Ihmisen ajatteluun liittyvien, moniulotteisten ja usein abstraktien ilmiöiden mittaus ei onnistu ihan suoraviivaisesti. Väliin tarvitaan vaihe, jota kutsutaan *operationalisoinniksi*. Tämä hankalan kuuloinen sana tarkoittaa sitä, että tutkittavan ilmiön ulottuvuudet, siis käsitykset ja asenteet, puretaan pienemmiksi paloiksi ja lopulta konkreettisiksi kysymyksiksi tai väitteiksi, jotka voidaan esittää vastaajille. Operationalisointi on useimmiten välttämätön vaihe myös kieleen liittyvissä kyselytutkimuksissa. Esimerkiksi tunnesanojen merkityksiä tai vihapuheeseen liittyviä asenteita ei yleensä voida kysyä suoraan. Tutkijalla voi olla hypoteesi, jonka mukaan tietyssä ihmisryhmässä ei periaatteessa hyväksytä vihapuhetta mutta käytännössä itse tuotetaan sitä. On mietittävä, millaisten kysymysten välityksellä ilmiöstä saadaan tietoa. Yhden abstraktin asian mittaus voi vaatia monta kysymystä. Näistä kysymyksistä syntyy mahdollisesti kokonainen mittari, joka puolestaan muodostaa osan kyselylomaketta.

Artikkelin aluksi tarkastelemme kyselyn historiallista ja filosofista taustaa, kielentutkimuksessa tehtyjä kyselyjä sekä kyselyyn liittyviä eettisiä näkökulmia (luku 2). Sen jälkeen esittelemme kyselyn käytännön vaiheita (luku 3). Päätösluvussa (4) kokoamme artikkelimme tärkeimmät näkökohdat.

# 2. Kyselyn taustaa

## 2.1. Kyselyn synty ja kehitys

Toisin kuin monet muut kielentutkimuksessa käytettävät menetelmät kysely on saanut alkunsa ja kehittynyt tieteelliseksi tutkimuskeinoksi kielitieteen ulkopuolella. Tiedon keruun ja analysoinnin menetelmiä kehitettiin tilastotieteessä, kun taas mittauksen haasteisiin perehdyttiin pitkään enemmän psykologiassa.

Ensimmäiset dokumentoidut kyselytutkimukset juontavat juurensa väestölaskentaan ja markkinatutkimuksiin. Vuonna 1916 *The Literary Digest* -viikkolehti teki kyselyn lähettämällä miljoonia postikortteja ympäri Yhdysvaltoja ja ennusti tulosten perusteella oikein Wilsonin valinnan presidentiksi. Tutkimus onnistui kerrasta toiseen, kunnes vuonna 1936 tapahtui kömmähdys. Tutkijat eivät ottaneet huomioon sitä, että heidän 2,3 miljoonaisesta vastaajajoukostaan valtaosa oli republikaaneja. Samaan aikaan George Gallup ennusti oikein Rooseveltin äänivyöryn tekemällä huomattavasti pienemmän mutta väestötasolla edustavan kyselytutkimuksen (ks. Wikipedia s.v. opinion poll).

Kyselyn avulla saadun aineiston luotettavuudessa ei kuitenkaan ole kyse yksinomaan luotettavasta tiedon keruusta. Riippumatta siitä, keille kysymykset esitetään, ne pitää myös esittää pätevällä tavalla. Kysely on keino mitata monimutkaisiakin ilmiöitä. Mittauksen haasteisiin ihmistieteissä perehdyttiin alun alkaen psykologiassa, varhaisten älykkyystutkimusten yhteydessä (Spearman 1904). Sittemmin niin psykologit, kasvatustieteilijät, yhteiskuntatieteilijät kuin tilastotieteilijät ovat kehittäneet kyselytutkimuksessa tarvittavia menetelmiä sekä tiedon keruun ja mittauksen että analysoinnin tarpeisiin. (Kyselymenetelmän käytöstä ja kehityksestä Suomessa ks. Alastalo 2005.)

## 2.2. Millaisissa kielitieteen tutkimuksissa kyselyä on käytetty

Ensimmäiset dokumentoidut kielitieteen kyselyt löytyvät 1800-luvun lopun murretutkimuksista, joissa kartoitettiin Saksassa ja Ranskassa puhuttuja kielimuotoja (Chambers & Trudgill 1980, 15–17). Postikyselyt vaihtuivat pian koulutettujen murteenkerääjien kenttämatkoihin. Kyselyt olivat osa myös suomalaisen dialektologian varhaisia vaiheita (ks. Haastattelu aineistonkeruun metodina sosiolingvistiikassa tässä kirjassa).

Kysely tutkijan ulkopuolelle laajennettuna kielitajuna on ollut käytössä jo paljon kauemmin. Kyselyn taustalla on yleensä tutkijan kielitajun hyödyntäminen (ks. Kielitaju kielentutkijan työkaluna tk.). Kysely on lähellä myös empiirisen kielentutkimuksen menetelmiä, etenkin haastattelua (ks. Haastattelu aineistonkeruun metodina sosiolingvistiikassa tk.) ja kenttätyömenetelmiä (ks. Kielten dokumentointi ja kieliopin kuvaus tk.). Läheisyys haastatteluun näkyy myös terminologiassa: englannin *survey* kattaa sekä kysely- että haastattelututkimuksen, eikä sanalle ole vakiintunutta suomennosta.

Varhaisten vaiheiden jälkeen kyselyjä tai kyselyn tapaisia menetelmiä on käytetty eri puolilla maailmaa kielen rakenteiden, merkitysten ja käytön tutkimuksessa (esim. Schilling 2013). Kyselyllä on mitattu mm. kielitaitoa ja asenteita ja kartoitettu uhanalaisten kielten tilannetta. Suomalaisessa kielitieteessä kyselyä on käytetty etenkin sosiolingvistiikassa ja dialektologian tutkimuksessa. Sillä on tutkittu mm. murrematkimuksia (Leino 1968), murreasenteita (esim. Mielikäinen & Palander 2002), Helsingin puhekieltä (esim. Ainiala & Lappalainen 2010) ja henkilöviittauksia (Priiki 2011; Siitonen 2016). Kysely on ollut melko suosittu menetelmä myös leksikaalisessa semantiikassa, jossa tutkimuskohteina ovat olleet mm. onomatopoeettisten sanojen tuntemus ja tulkinta (Leskinen 1991), polysemia (Raukko 2002) ja synonymia (Jürvetson 2015; Vanhatalo 2002; 2005). Sitä on käytetty myös kielitieteen terminologian tutkimuksessa (Luodonpää-Manni 2013).

Kielitieteen kyselytutkimukset ovat perustuneet yleensä harkinnanvaraisella näytteellä kerättyyn aineistoon (ks. luku 3.3.), vaikka myös

otantaan perustuvia tutkimuksia on tehty. Hyvänä kotimaisena esimerkkinä otannalla kerätystä aineistosta on englannin kielen asemaa 2000-luvun Suomessa selvittänyt kyselytutkimus, johon saatiin Väestörekisterikeskuksen kautta noin 1 500 vastaajaa (Leppänen ja muut 2011).

Kielentutkimuksessa käytettyjen kyselyjen metodologinen ja terminologinen johdonmukaisuus on vaihdellut. Tieteenalojen välisen keskustelun lisääntyessä 2000-luvulla on tietoisuus muilla aloilla kehittyneistä metodeista lisääntynyt ja vaatimus metodologisesta tarkkuudesta kasvanut. Täsmällisyyttä edistävät osaltaan myös tieteenalojen väliset julkaisukanavat, jotka vaativat yhdenmukaisia metodologisia käytäntöjä eri tieteenalojen kyselytutkimuksissa.

## 2.3. Kyselyn käytön perusteet

Kysely on edelleen tärkeä tutkimusmenetelmä yhteiskunta- ja käyttäytymistieteissä. Menetelmä on altis sekä otanta- että mittausvirheille, mutta sille on vaikea keksiä vaihtoehtojakaan. Yksi kyselymenetelmän perusajatuksia on, että tunnetut virheet pyritään tiedostamaan, ottamaan huomioon ja korjaamaan. Usko signaalin eli tietoa välittävän ilmiön löytymiseen säilyy silloinkin, kun kaikkia häiriötekijöitä ei tunneta. Joskus on spekuloitu sillä, voisivatko esimerkiksi neurologiset mittaukset tehdä turhiksi kyselyjä. Toisaalta myös niissä on häiriötekijöitä, eikä virheetöntä mittausta olekaan (Vehkalahti 2019, 17).

Kielitieteessä oletetaan, että kyselyllä voidaan riittävällä tarkkuudella tavoittaa vastaajan kielitaju tai ainakin vastaajan käsitys kielitajustaan (ks. Kielitaju kielentutkijan työkaluna tk.). Kyselyllä voidaan tutkia vastaajan tietoisten käsitysten lisäksi myös tiedostamattomia asioita, kuten asenteita tai kielen käyttöä. Kysely tulee mukaan tutkimusasetelmaan usein silloin, kun jotain kielen ilmiötä on vaikeaa tai mahdotonta tutkia yksinomaan kielitajulla (ks. Kielitaju kielentutkijan työkaluna tk.) tai havainnoivilla menetelmillä (ks. Johdannon luku 1 tk.).

Tieteenfilosofisesti kiinnostava on kysymys kyselymenetelmän muokattavuudesta kielentutkimuksen tarpeisiin tai tutkijan rajoituksiin. Voiko ja saako kyselyn käyttö kielentutkimuksessa poiketa niistä

metodologisista normeista, jotka vallitsevat kielitieteen ulkopuolella? Olisiko tämä joissain tapauksissa jopa suotavaa? Millaisia nämä mahdolliset normista poikkeamiset voisivat olla, jotta menetelmää vielä voidaan nimittää *kyselyksi*? Kielentutkija voi kohdata työssään tämäntapaisia metametodologisia kysymyksiä. Perehtyminen kyselymenetelmän metodologisiin normeihin muilla tieteenaloilla voi auttaa vastausten löytämisessä (esim. Croves, Fowler, Couper, Lepkowski, Singer & Tourangeau 2009).

## 2.4. Kyselyyn erityisesti liittyviä eettisiä seikkoja

Kyselyyn liittyy ilmeisiä eettisiä kysymyksiä, esimerkiksi vastaajien henkilötietojen keruu ja säilytys, vastaajien mahdollinen alaikäisyys sekä tutkimuskohteena olevien aiheiden mahdollinen arkaluontoisuus. Kyselyaineiston keruu voi tarvita keruupaikasta (esim. oppilaitokselta) tutkimusluvan. Kyselyaineistojen keruuta ja säilytystä koskevat eettiset säännökset voivat vaihdella eri maissa, mikä tulee ottaa huomioon kansainvälisiin julkaisuihin tähtäävissä tutkimuksissa (ks. Johdannon luku 4 tk.).

Vähemmän ilmeisiä mutta tärkeitä eettisiä kysymyksiä ovat metodologinen ja terminologinen systemaattisuus sekä tulosten tulkinta, jossa ongelmat liittyvät usein näyteaineistoon perustuvien tulosten yleistämiseen vastaajajoukon ulkopuolelle (tulkinnasta lisää ks. luku 3.3.).

## 2.5. Terminologisia haasteita

Kyselymenetelmän systemaattinen kehittäminen yhtäällä (tilastotiede, psykologia) ja spontaani kehittyminen toisaalla (esim. kielentutkimus) on aiheuttanut runsaasti terminologista variaatiota. Ilmeisen tahattomalla variaatiolla voi olla haitallisia seurauksia. Virheellinen terminologia voi aiheuttaa metodologisia virheitä itse tutkimusprosessissa. Esimerkiksi aineistonkeruuseen liittyvien termien *otos* ja *näyte* sekoittaminen voi johtaa tulosten virheelliseen yleistämiseen (määritelmistä

ks. luku 3.3.). Terminologinen vaihtelu voi myös estää muiden alojen tutkijoita löytämästä hakusanoilla, mitä kielentutkimuksessa on tehty. Itse *kysely*-termi on monissa tutkimuksissa nimetty *testiksi*, esim. *kyselytesti* (Raukko 2002; Vanhatalo 2002; Vanhatalo 2005; mahdollisesti näitä seuraten myös Itkonen & Pajunen 2010 sekä Luodonpää-Manni 2013), *parafraasitesti* (Västi 2012) ja *kansandialektologinen testi* (Räsänen & Palander 2015). *Kysely*-termissä pitäytymistäkin sentään esiintyy (esim. Priiki 2011). Sanan *testi* käyttäminen antaa turhan kapean kuvan tilastollisen analyysin mahdollisuuksista: tutkimuksessa saatetaan olla kiinnostuneita tiivistämään ja kuvailemaan tutkitun ilmiön ominaispiirteitä myös ilman, että analyyseihin liittyy mitään tilastollisia testejä. Kysely voi siis olla esimerkiksi alustava kartoitus jostain ennen tutkimattomasta ilmiöstä. Kyselyn tuottama tieto taas voi nostaa esiin hypoteeseja, joita voidaan myöhemmin ehkä myös testata, joko samalla tai kokonaan uudella aineistolla.

# 3. Kuinka kysely tehdään

Käymme tässä luvussa pääpiirteittäin läpi kyselyn tärkeimmät vaiheet tutkimuksen suunnittelusta (3.1.) kyselylomakkeen laatimiseen (3.2.), aineiston keruuseen (3.3.), analyysiin (3.4.) ja tulosten tulkintaan (3.5.). Keskitymme erityisesti kyselylomakkeen suunnitteluun.

Metodioppaissa esitetään usein tapaustutkimuksia, joissa kerrotaan, miten ideasta päästiin lopputulokseen. Metodologisten pohdintojen ja harha-askelten siivoaminen tutkimusraporteista voi kuitenkin luoda sen käsityksen, etteivät tällaiset kuuluisi tutkimusprosessiin. Tätä virheellistä käsitystä oikaistaksemme asetumme tässä luvussa lukijan vierelle tekemään aitoa kielitieteen alan tutkimushanketta (Vanhatalo ja muut 2019). Olemme havainnoineet tutkimusprosessiamme reaaliaikaisesti ja kirjoittaneet sitä auki.

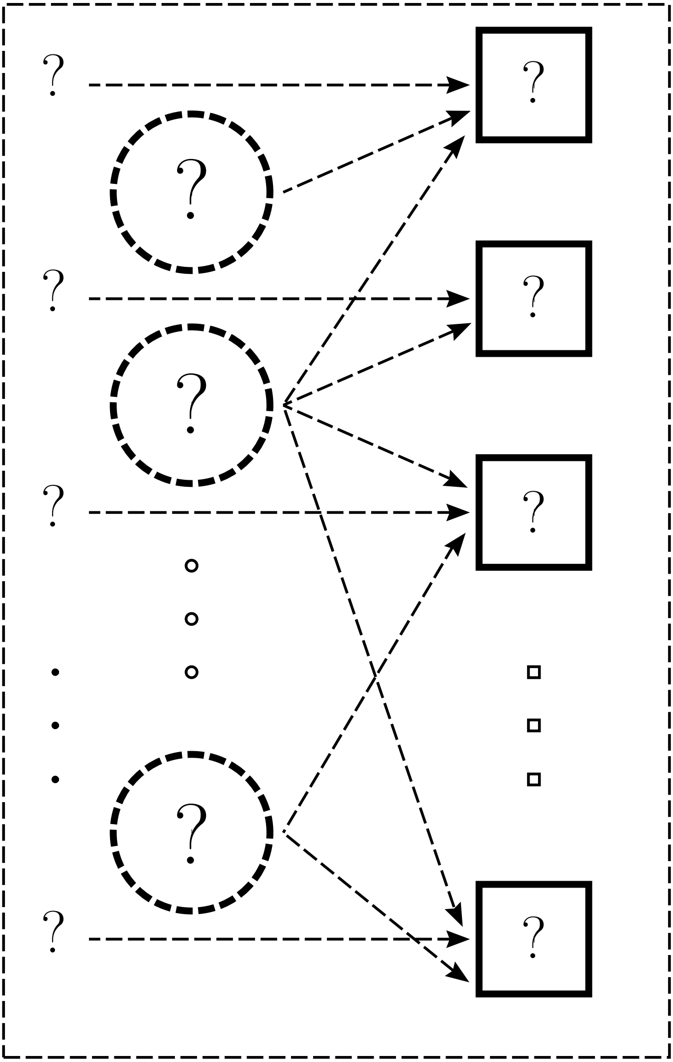
Tapaustutkimuksemme kuuluu leksikaalisen semantiikan eli sanojen merkityksen tutkimuksen alaan. Se vastaa laajuudeltaan suunnilleen pro gradu -työtä. Teoreettisena viitekehyksenä on alkusanakieli (engl. *Natural Semantic Metalanguage*), jossa semanttisia sisältöjä muotoillaan yksinkertaisemmin uudestaan (Goddard & Wierzbicka 2014; Vanhatalo & Tissari 2017). Tapaustutkimuksemme tavoite on selvittää, tunnistavatko äidinkieliset suomen puhujat alkusanakielisiä tunnesanojen (*ilo, suru, inho, viha, pelko, hämmästys, rakkaus*) selityksiä. Tutkimusasetelmassa etsitään myös lisätietoa tunnesanojen mahdollisesta universaaliudesta.

## 3.1. Tutkimuksen suunnittelu

Kyselytutkimuksen suunnittelun pohjalla ovat jo luvussa 1 esitetyt isot tutkimuskysymykseen liittyvät aiheet. Vielä tätä varhaisempaan suunnitteluvaiheeseen on liittynyt kysymys tutkimusmetodien valinnasta. Kyselyyn päädytään usein silloin, kun vastausten saaminen tutkimuskysymyksiin on muilla menetelmillä vaikeaa tai mahdotonta. Samalla pidetään todennäköisenä, että tutkittava aihe on riittävän mielenkiintoinen houkuttelemaan vastaajia. Suunnitteluvaiheessa selvitetään myös se, yhdistetäänkö kyselyyn muita menetelmiä ja millaisia tutkimuseettisiä kysymyksiä aiheeseen mahdollisesti liittyy.

Kyselytutkimuksen käytännön suunnittelu kannattaa aloittaa tutustumalla kyselymenetelmän käyttäjän parhaaseen apuvälineeseen, mittausmalliin, joka periytyy jo varhaisista psykologisista tutkimuksista (Thurstone 1931; Spearman 1904). Mittausmalli auttaa hahmottamaan tutkimuksen kokonaisuutta ja helpottaa lomakkeen laatimista. Suosittelemme, että mittausmalli tulostetaan riittävän isolle paperille, sitä täydennetään kynällä ja se kiinnitetään työpöydän yläpuolelle. Mittausmalliin ei erikseen eksplikoida tutkimuskysymystä, mutta se kulkee läpi koko tutkimuksen punaisena lankana.

Kuvan 1 vasemman laidan katkoviivalla merkityt ympyrät ovat tutkijan alustavia ideoita ja havaintoja abstraktin ilmiön mahdollisista ulottuvuuksista. Kysymysmerkit kuvaavat tutkimuksen alkuun kuuluvaa




Kuva 1. Mittausmalli alkutekijöissään (Vehkalahti 2019, 21).

epävarmuutta. Tutkimuksen myöhemmissä vaiheissa vasemman laidan ympyröitä kutsutaan *faktoreiksi*. Kielentutkija lähtee tutkimaan moniulotteista ilmiötä, jossa on erilaisia ulottuvuuksia. Katkoviiva korostaa sitä, että kyse on tutkijan teoreettisista ajatusrakennelmista, joita ei kerrota tutkimukseen osallistuville aineiston keruuvaiheessa.

> Kirjoitamme vasemman laidan ympyröihin tunnesanat. Mittausmallimme kuvaa näin seitsenulotteista tunnesanojen kokonaisuutta: sanoja ja niiden oletettuja merkityssisältöjä.

Ympyröissä olevien sisältöjen vahvistamiseksi tai kumoamiseksi tarvitaan tietoa, jota saadaan oikean laidan isoihin laatikoihin muodostettavilla kysymyksillä tai väitteillä. Näitä kysymyksiä tai väitteitä kutsutaan *osioiksi*. Jotta ympyröissä oleviin oletuksiin päästään käsiksi, osioita täytyy joskus olla paljonkin. Kysymysten muodostamisen prosessia kutsutaan *operationalisoinniksi*. Vain tutkijan tiedossa olevat abstraktiot ovat kuvion vasemmassa laidassa, vastaajan näkemä konkretia eli kysymykset oikeassa.

> Oikean laidan laatikoihin kirjaamme sanojen merkityspiirteet ja kokonaiset merkitysten selitykset. Näistä syntyvät lomakkeemme kysymykset. Kun osa merkityspiirteistä liittyy useampaan tunnesanaan, tiedämmekö lopulta, mitä tunnetta milläkin kysymyksellä mitattiin? Tilastotieteilijä vakuuttaa, että kaaoksesta tulee löytymään järjestys.
>
> Osa merkityspiirteistä tulee asettumaan selvästi joidenkin tunnesanojen kohdalle. Näin esimerkiksi kaikki pelkoon liittyvät kysymykset korreloivat keskenään. Yksittäiset merkityspiirteet ja niiden yhdistelmät lisäävät luotettavuutta. Isoon kysymysmäärään sulautuvat myös sellaisenaan liian ilmeiset kokonaiset merkityksen selitykset.

Kuvassa isojen ympyröiden välissä näkyvät pikkuympyrät kuvaavat ilmiön sellaisia ulottuvuuksia tai tutkijan taustaoletuksia, joita tutkija tiedostamattaan tai tarkoituksella jättää tutkimuksen ulkopuolelle.

Pikkulaatikot puolestaan ovat mahdollisia kysymyksiä, joita tutkija muodostaa tai ei muodosta mutta jotka eivät tule kyselyyn. Molemmat pikkumerkinnät kuvaavat myös sitä, että ympyröiden tai laatikoiden oikea lukumäärä ei välttämättä heti ole tutkijallekaan selvää; ne tyypillisesti tarkentuvat suunnittelun edetessä.

Ihannetapauksessa yhdestä ympyrästä lähtee viivat moneen laatikkoon mutta jokaisesta laatikosta lähtee viiva yhteen ympyrään. Tämä tarkoittaa sitä, että jokaista tutkittavaa ulottuvuutta voidaan mitata monella kysymyksellä mutta jokaisella kysymyksellä mitataan lähtökohtaisesti vain yhtä ulottuvuutta.

Ihmistieteiden mittausasetelmissa on usein paljon epävarmuutta ja mittausvirheitä, jotka vähentävät tulosten luotettavuutta tai tarkkuutta eli reliabiliteettia. Tämä on osittain seurausta niiden tutkimuskysymysten laadusta, joihin ihmistutkimuksessa haetaan vastausta. Virheitä voidaan vähentää niin, että jokaista tutkittavaa oletusta kysytään moneen kertaan ja useammalta kantilta. Pieniksi paloitellut ja tutkimuskohdetta monipuolisesti mittaavat kysymykset lisäävät toistoa (redundanssi), joka puolestaan lisää tulosten luotettavuutta. Mittaus saadaan näin vankemmaksi, eivätkä tulokset ole kiinni yksittäisistä kysymyksistä. Reliabiliteetin rinnalla on muistettava huolehtia tutkimusasetelman ja kysymysten validiteetista eli siitä, mitataanko sitä, mitä on tarkoitus (ks. Johdannon luku 2 tk.).

Tutkimusasetelman suunnittelun ja lomakkeen kysymysten rakentamisen tulee olla kiinteässä yhteydessä tutkimuksen taustalla vaikuttavaan teoreettiseen viitekehykseen. Mikä tahansa tutkimusasetelma ei ole teoreettisesti validi, eikä mikä tahansa tutkimuskohteeseen jotenkuten liittyvä kysymys ole tieteellisesti toimiva mittari. Tieteelliseen ajatteluun (ks. Johdannon luku 2 tk.) sitoutunut tutkimusasetelma eroaa iltapäivälehtien gallupeista, vaikka niissäkin kerätään ja tarkastellaan kyselyaineistoa. Jostain syystä juuri kyselytutkimusten kohdalla on kiinnitetty huomiota heikkoon keskusteluyhteyteen muun tutkimuksen kanssa (Hirsjärvi, Remes & Sajavaara 2009, 195). Mitään periaatteellista syytä pinnallisuudelle ei pitäisi olla; kysely menetelmänä on yhdisteltävissä hyvin monenlaisiin teoreettisiin viitekehyksiin. Voi tosin olla, että aineiston keruu ja mittaaminen vievät kyselyissä enemmän aikaa kuin

muiden menetelmien kohdalla, eikä tutkijalla riitä aikaa analyysin syventämiseen tai teoreettisiin pohdintoihin. Voi myös olla helppoa innostua jostain kielen ilmiöstä, tehdä ad hoc -kysely ja saada hauskoja tuloksia. Ilman vankkaa taustateoriaa tuloksilla on kuitenkin riski jäädä ohuiksi, kertaluontoisiksi ja kielitieteellisen keskustelun ulkopuolelle (ks. Johdannon luku 1 tk.).

> Tutkimuksemme on validi: mittarimme mittaa sitä, mitä on tarkoituskin. Tutkittavien tunnesanojen valikoima tulee aiemmasta psykologisesta tutkimuksesta ja lomakkeen kysymykset hieman muokattuina kielentutkimuksesta. Tuloksemme keskustelevat suoraan aiemman alkusanakielestä tehdyn tunnesanatutkimuksen kanssa.

Kandidaatintutkielmaa tai pro gradua tekevän on jo ajankäytöllisesti järkevää sitoa opinnäytteensä vakiintuneeseen teoriaan tai tutkimussuuntaukseen, jonka pohjalle kysely rakennetaan. Kannattaa harkita, lähteekö tutkimaan sellaista, mitä kukaan ei ole ennen tutkinut. Tutkimuskirjallisuudesta voi löytyä valmiita johonkin teoriaan sitoutuneita kyselylomakkeita, joita hyödyntää jopa sellaisenaan. On huomattava, että joskus jossain ajassa ja paikassa hienosti toiminut kysely ei ehkä enää toimikaan. Aikanaan hyvinkin käytetty tutkimusmittari vaatii ennen pitkää päivittämistä. Siinä missä kandin tai gradun tekijä välttää metodologisia kokeiluja, väitöskirjantekijä keksii uuden mittarin tai korjaa vanhaa. Tutkimusprosessi keskustelee silloin aiemman tutkimuksen kanssa myös metodologian tasolla.

Tutkimuksen suunnitteluvaiheessa pyritään ennakoimaan ja kuvittelemaan tutkimuksen eri vaiheet. Tutkimuskysymys määrää suunnan sille, mitä ollaan tekemässä. Kaikkea ei kuitenkaan voi ennakoida. Tämä on odotuksenmukaista yllätyksellisyyttä ja osa tutkimusprosessin kulkua. Prosessin kuluessa aineisto saa tehdä oman osuutensa. Joskus tutkimuskysymystä joudutaan tarkentamaan ja jopa muuttamaan kesken tutkimusprosessin. Matkan varrella voi herätä jokin metodologiaan tai analyysiin liittyvä idea, jolle kehitellään jälkikäteen tutkimuskysymys. Tällaisessa ei sinänsä ole mitään tuomittavaa. On päinvastoin tylsää ja

jopa vaarallista tehdä niin etukäteen määrättyä tutkimusta, että kaikki hypoteesit on päätetty etukäteen ja aineisto kerätään vain vahvistamaan hypoteesia.

## 3.2. Kyselylomakkeen laatiminen

Kyselylomakkeen laatiminen on kyselyn tärkein vaihe. Tarkastelemme tässä luvussa lomakkeen laatimiseen liittyviä taustatekijöitä, yksittäisten kysymysten sisältöä ja rakennetta, kysymysten asettelua lomakkeessa sekä vastausohjeita.

Tutkijan omat ennakkokäsitykset ja kielitaju vaikuttavat aina siihen, miten tutkimushypoteesit muodostetaan ja millaisia kysymyksiä lomakkeeseen tehdään (Itkonen & Pajunen 2010; Kielitaju kielentutkijan työkaluna tk.). Kyselykään ei siis koskaan ole neutraali ja objektiivinen tutkimusasetelma (ks. Johdannon luku 2 tk.). Tutkijan tulee pyrkiä tiedostamaan tämä niin lomaketta laatiessaan kuin myös tuloksia kirjatessaan. Tässä auttaa avoimuus kriittiselle keskustelulle sekä tutkimuksen huolellinen valmistelu.

Prosessiin liittyy erilaisia häiriötekijöitä myös vastaajan puolelta. Vastaaja ei välttämättä keskity vain kyselyyn vaan tekee samalla muutakin: vastaa puhelimeen, juttelee muiden kanssa tai selaa nettiä – ja ehkäpä juuri kyselyyn liittyviä sivuja. Tutkija voi onneksi varautua myös tähän. Kysymysten pitää olla lyhyitä ja helppoja, ja niitä pitää olla riittävän paljon (ks. Tieteen termipankki s.v. redundanssi). Vastaaja voi myös haluta näyttää fiksulta ja alkaa pohtia, mitä hänen odotetaan vastaavan (ks. Tieteen termipankki s.v. sosiaalisen suotavuuden vastaustaipumus), kun taas tutkija tarvitsee vain rehellisen vastauksen. On siis hyvä, jos lukuisat yllättävistä kulmista asiaa lähestyvät kysymykset "sekoittavat" vastaajan. Vastaajan ei tarvitse ymmärtää tutkimuksen kokonaisuutta, sillä se on tutkijan tehtävä. Vastaajan tulee vain vastata kuhunkin kysymykseen.

Yksittäisten kysymysten sisältö ja rakenne määräytyvät tutkimusaiheen mukaan. Mittausmallin oikean puolen laatikot kertovat, mitä vastaajilta on kysyttävä, jotta saadaan vastauksia tutkimuskysymyksiin. Kysymysten on oltava paitsi lyhyitä myös yksiselitteisiä, eivätkä ne saa

johdatella vastaajaa. Kysymysten strukturoituneisuuden aste ja vastausvaihtoehdot (ml. täytevaihtoehdot, Likert-asteikot) vaikuttavat suoraan siihen, miten tuloksia voidaan analysoida. Perussääntönä voidaan ehkä pitää sitä, että avoimia vastauksia tulee välttää, jos aineistoa tullaan analysoimaan tilastollisilla menetelmillä. Toisaalta on muistettava, että avoimet vastaukset voivat tuottaa erittäin kiinnostavaa laadullista aineistoa.

Kaikki kysymykset eivät ole ehkä yhtä tärkeitä tutkijan näkökulmasta. Tärkeät kysymykset vastauksineen voidaan jälkikäteen tarvittaessa poimia erikseen tarkasteltavaksi. Lomakkeessa voi olla yhteensä kymmeniä lyhyitä kysymyksiä tai väittämiä, joten turhat kysymykset olisi hyvä jättää jo etukäteen pois. Tässä voi olennaisesti auttaa pilotti, jota käsittelemme tuonnempana. Vastaajien taustatiedot (esim. äidinkieli tai ikä) ovat tärkeitä tutkimuksen tulosten esittelyssä ja yleistämisessä (ks. luku 3.5.). Tutkija voi myös haluta analysoida tuloksia joidenkin taustatietojen mukaan. Kysymyksistä ja niiden muotoilusta kannattaa keskustella seminaariryhmässä tai muiden tutkijoiden kanssa.

Tutkimuksen esittely vastaajille ja vastausohjeet tulee miettiä huolellisesti. Tutkimusaiheen esittelyn tulee olla lyhyt ja yleistasoinen. Vastaaja olisi kuitenkin saatava vakuuttumaan siitä, että kyselyyn kannattaa osallistua ja tehdä se loppuun. Alussa on luvattava, että vastaamisen voi halutessaan jättää kesken, ja lopussa tavalla tai toisella saatava vastaajan suostumus aineiston käyttöön tutkimustarkoituksessa (engl. *consent form*). Vastausohjeiden on oltava lyhyitä ja täysin yksiselitteisiä. Sekä esittelytekstistä että ohjeista kannattaa pyytää palautetta pilottiin osallistuvilta.

Huomioon on otettava myös kysymysten asettelu (yksi vai useampi sivulle), järjestys lomakkeessa (mihin on helppoa vastata aluksi, vaikuttavatko aiemmat kysymykset tulevien vastauksiin), esitysjärjestyksen satunnaistaminen (voi olla niin hyötyä kuin haittaa) ja lomakkeen pituus (kauanko vastaajalla menee aikaa – tässä on syytä olla realistinen, jottei aliarvioi vastaamiseen kuluvaa aikaa!). Sijoittamalla taustatietokysymykset sisältökysymysten jälkeen kyselyn loppuun vältetään kuulustelumaisuus lomakkeen alkupäässä. Vastaajan kannalta on hyvä, jos lomakkeessa näkyy kyselyn kulku ja jäljellä olevien kysymysten tai sivujen määrä.

Kaikki kysymykset ja ohjeet kannattaa suunnitella teksti- tai taulukkotiedostossa, eikä lomakeohjelmaa tule ottaa käyttöön ennen kuin kaikki kysymykset tuntuvat valmiilta. Kysymysten muokkaaminen lomakeohjelmassa voi olla hyvin hidasta ja hankalaa. Täysimittainen sähköinen lomake pilottitutkimusta varten kannattaa tehdä vasta sitten, kun on varma kysymysten muotoiluista ja järjestyksestä.

Lomakeohjelman valinnassa kannattaa ottaa huomioon paitsi lomakkeen tekniset ominaisuudet (etenkin toimivuus eri laitteilla ja selaimilla) myös käyttäjätuki. Lomakeohjelman automaattiasetuksista voi olla hyötyä tai haittaa selkeän lomakkeen aikaansaamisessa. Oman instituution käyttämä lomakeohjelma voi puutteistaan huolimatta olla järkevä valinta, jos tarjolla on käyttäjätukea niin suunnittelussa kuin myös mahdollisen katastrofin sattuessa.

Kun lomake on tutkijan mielestä lopulta valmis, on aika tehdä esipilotti. Esipilotissa muutama tutkijan valitsema henkilö täyttää lomakkeen verkossa ja raportoi havainnoistaan. Lomaketta testataan myös erilaisilla etenkin kannettavilla päätelaitteilla. Kun esipilotissa mahdollisesti ilmenneet ongelmat on korjattu, voidaan siirtyä varsinaiseen pilottiin. Sen tehtävänä on näyttää, toimiiko tutkimusasetelma aineistonkeruusta aina analyysiin asti niin kuin on ajateltu. Pilottiin riittää tyypillisesti 10–15 vastaajaa, jotka parhaassa tapauksessa vastaavat taustaprofiililtaan varsinaiseen tutkimukseen osallistuvia. Pilottiin osallistuvilta kannattaa myös kysyä kokemuksia lomakkeesta. Pilotin jälkeen voidaan vielä tehdä muutoksia lomakkeeseen, mutta kun varsinainen aineistonkeruu alkaa, muutosten tekeminen ei enää ole mahdollista.

> Alussa tahdomme tehdä mahdollisimman lyhyen kyselyn ja yksinkertaisia kysymyksiä. Näin suunnittelemme:
>
> ”En tiedä, mitä kohta tapahtuu. Voi tapahtua pahoja asioita. En tiedä, mitä voin tehdä. Se tuntuu minusta pahalta.”
>
> Mistä tunteesta on mielestäsi kyse? Kerro vastauksesi lyhyesti.
>
> ---
>
> (Odotamme vastaukseksi sanaa *pelko*.)

Tilastotieteilijä kyseenalaistaa suunnitelmamme liian yksinkertaisena ja häiriötekijöille alttiina. Asetelma ei ole riittävän luotettava, koska siitä puuttuu redundanssia. Lisäksi avointen vastausten analyysi on hankalaa. Lisäämme kysymysten määrää ottamalla mukaan merkitysten selitysten osakomponentit. Luovumme avoimista vastauksista ja annamme vastaajalle listan, josta ruksia sopivat tunnesanat. Päätämme hämätä vastaajaa laittamalla listalle myös seitsemän täytesanaa. Pilottipalautteen jälkeen lisäämme listan loppuun myös vaihtoehdot *en osaa sanoa* ja *jokin muu, mikä?*

Näin teemme:

”minusta tuntuu pahalta”

Mihin tunteeseen tämä mielestäsi liittyy? Merkitse rastilla yksi tai useampi vaihtoehto.

□ ahdistus
□ hämmästys
□ häpeä
□ ilo
□ inho
□ kateus
□ onnellisuus
□ pelko
□ rakkaus
□ riemu
□ suru
□ tyytyväisyys
□ viha
□ ylpeys
□ en osaa sanoa
□ jokin muu, mikä? ___________

Jokaista tunnesanaa mittaa nyt monta kysymystä. Toisaalta osa kysymyksistä mittaa nyt useampaa tunnesanaa, kun esimerkiksi *minusta tuntuu pahalta* kuuluu niin surun kuin vihan selityksiin. Vastauksiksi saamme lähes pelkkiä rasteja. Hylkäämme ajatuksen merkityskomponenttien intensiteetin mittauksista liukuvalla asteikolla liian mutkikkaana analysoitavana.

33 monivalintakysymyksen jälkeen teemme vielä muutaman käännöskysymyksen: *Miten sanoisit suomeksi: I'm happy. / I'm angry* jne. Taustatiedoista kysymme syntymävuotta, sukupuolta, ensikieltä ja mahdollisia kieliaineiden tai psykologian opintoja ja sitä, minkä kurssin yhteydessä kyselyyn on osallistuttu.

Tavoitteemme on saada vastaaja vastaamaan nopeasti sen, mitä mieleen tulee.

Haluamme kysymykset satunnaiseen järjestykseen, mutta emme halua kokonaisia selityksiä aivan lomakkeen alkuun. Käyttämämme lomakeohjelma ei taivu kysymysten satunnaiseen esitysjärjestykseen, joten huolehdimme siitä erillisellä ohjelmalla. Kysymysten muokkausvaiheessa teemme myös ensimmäisen esipilottilomakkeen e-lomakepohjaan. Lyhyt esipilottilomake sisältää mallisivut kaikista kysymystyypeistä. Testailemme minilomaketta eri tietokoneilla ja kännyköillä.

Tässä luvussa olemme tarkastelleet lomakkeen laatimista. Lomakkeen suunnittelu ei tosiasiassa kuitenkaan rajoitu vain lukuun 3.2., vaan koskee mitä suurimmassa määrin myös analyysia (luku 3.4.). Suunnitteluun vaikuttavat myös julkaisukanavan asettamat rajoitukset esimerkiksi merkkimäärälle, taulukoille ja kuville.

Lomakkeen suunnittelun tärkeyttä ei voi liiaksi korostaa. Kun mittaukset on lopulta tehty, tutkijalla on joukko numeroita (ja mahdollisia avoimiin kysymyksiin tulleita vapaamuotoisia vastauksia). Koko tutkimus riippuu tämän jälkeen näistä numeroista, joita ei voi enää uusia ilman uutta mittausta. Mikään analyysimenetelmä ei pelasta huonosti

tehtyä mittausta. Tämä on koko kyselymenetelmän käytön tärkeimpiä huomioita.

> Tapaustutkimuksemme pilotissa on 16 vastaajaa. Tulosten alustava analyysi näyttää hyvältä. Pilotin perusteella lisäämme mahdollisuuden antaa myös avoin *muu, mikä?* -vastaus rastien lisäksi. Sovimme kuitenkin, ettemme analysoi tuon kysymyksen vastauksia tätä tutkimusta varten.[1]

## 3.3. Aineiston keruu eli vastaajien tavoittaminen

Aineiston keruuseen liittyy kaksi pääkysymystä, joihin kielentutkijan on itselleen vastattava: 1) kenen käsityksiä kielestä haluan tutkia ja 2) miten tavoitan nämä ihmiset. Ensimmäinen kysymys liittyy suoraan tutkimusten tulosten yleistettävyyteen, toinen tutkimuksen toteuttamiseen käytännössä. Tarkastelemme tässä luvussa näitä kysymyksiä.

Kyselytutkimuksen tekijän on yleensä jollakin tavalla rajattava kohderyhmänsä. Vain hyvin pienten kielten tutkijat voivat tehdä kokonaistutkimuksen ja ottaa kohderyhmäänsä kaikki tutkimansa kielen puhujat. Vastaajajoukon valintaan on kaksi päätapaa: satunnaisotanta ja harkinnanvarainen näyte. Otanta on tilastotieteen näkökulmasta (verrattain harvinaisen kokonaistutkimuksen lisäksi) tärkein aineiston keruutapa, koska siinä tulokset on mahdollista yleistää laajempaan kohdejoukkoon. Käytännössä moni kielitieteen tutkimus perustuu kuitenkin harkinnanvaraiseen näytteeseen eli jostakin (paremmin tai huonommin) perustellusti valitusta ihmisryhmästä kerättyyn aineistoon. Vastaajajoukon voi muodostaa esimerkiksi jokin opiskelijaryhmä.

Aineiston keruutavalla on suuri merkitys kyselyn tulosten tulkinnassa, ja siksi tutkijan on osattava tehdä selvä ero satunnaisotannan ja harkinnanvaraisen näytteen välillä. Satunnaisotannassa kaikilla perusjoukkoon kuuluvilla on samanlaiset mahdollisuudet päätyä vastaajaksi. Tulokset ovat tällöin suoraan yleistettävissä perusjoukkoon, josta otos on alun perin poimittu ja jota otos edustaa. Harkinnanvaraiseen

näytteeseen on jollain perusteella valittu tietty joukko vastaajia, eikä tuloksia lähtökohtaisesti pystytä yleistämään vastaajajoukon ulkopuolelle.

Kun esimerkiksi suomen kielen tutkija haluaa saada kaikkiin Suomessa asuviin äidinkielisiin suomen puhujiin yleistettäviä tuloksia, hän pyytää Väestörekisterikeskuksesta 2 000 hengen satunnaisotannan äidinkielenään suomea puhuvista kansalaisista. 2 000 vastaajaa on Suomessa tyypillinen määrä, jossa kustannukset eivät nouse liian suuriksi ja otoskoko riittää päteviin tuloksiin. Tutkija lähestyy kohdejoukkoaan kyselyllä, hankkii vastaukset ja analysoi tulokset. Satunnaisotantaan liittyy kuitenkin epävarmuustekijöitä. Väestörekisterikeskus ei välttämättä näe perusteita tietojen luovuttamiseen kyseistä tutkimusta varten. Vaikka yhteystiedot saisikin, vastausprosentti voi jäädä pieneksi, kato voi kasvaa suureksi ja korjaava aineistonkeruu osoittautua työlääksi. Jos samaiselle tutkijalle riittää tulosten yleistäminen esimerkiksi Helsingin yliopiston humanistisen tiedekunnan opiskelijoihin, hän voi pyytää tiedekunnan kaikkien opiskelijoiden yhteystietoja, tehdä näistä satunnaisotannan ja pyytää valittuja osallistumaan tutkimukseen. Ongelmat ovat samat kuin edellä mutta pienempinä. (Laaksonen 2013; Vehkalahti 2019.)

Otannan sijaan tutkija voi päättää kerätä harkinnanvaraisen näytteen, johon kuuluu jokin joukko ihmisiä, esimerkiksi tietyn luentokurssin osallistujat. Näyteaineiston keräämiseen päädytään usein silloin, kun tavoitteena on saada tutkijan kielitajua vahvistavaa alustavaa käsitystä tutkittavasta ilmiöstä. Näyteaineistoa voivat puoltaa aikataulusyyt: näyte on usein huomattavasti nopeampi kerätä kuin satunnaisotanta. Näyteaineisto mahdollistaa myös kontrolloidut kyselyolosuhteet esimerkiksi luokkahuoneessa. Näytetutkimus voi toimia pilottina myöhemmin tehtävälle satunnaisotantaan perustuvalle tutkimukselle, jolloin satunnaisotanta kerätään vasta sitten, kun näytteellä on saatu riittävän lupaavia tuloksia (ns. lumipallonäytteenotto). Näyteaineistoon saadaan uskottavuutta ja syvyyttä, kun se kohdennetaan johonkin ryhmään ja riittävillä taustakysymyksillä varmistetaan, että vastaajat kuuluvat riittävän vahvasti ko. ryhmään. Esimerkiksi jos vastaajan äidinkieli on olennainen taustatekijä, tätä on kysyttävä lomakkeessa ja analyysivaiheessa muiden kielten puhujat on poistettava aineistosta. (Laaksonen 2013; Vehkalahti 2019.)

Kielentutkimuksessa ei välttämättä ole aina ymmärretty aineistonkeruutavan ja yleistettävyyden kiinteää yhteyttä, ja yleistämään pyrkiviä tutkimuksia on tehty näyteaineistojen perusteella. Esimerkiksi Tummers, Heylen & Geeraerts (2005, 229–230) esittävät kyselyn (engl. *survey*) kollektiivisen introspektion muotona. He kiinnittävät huomiota observoijan paradoksiin ja kyselyn kontekstisidonnaisuuteen mutta eivät puhu mitään otoksesta, näytteestä tai yleistämisen ongelmista. Empiirisesti kerätyn aineiston semanttisia muuttujia mainitaan kyllä analysoitavan impressionistisesti, ja todellisten määrällisten menetelmien todetaan puuttuvan (mts. 248–249). Samoihin aikoihin ilmestyneessä Pohjois-Amerikan englannin fonologisessa atlaksessa (Labov, Ash & Boberg 2006) aineistonkeruuseen ja sen raportointiin on paneuduttu huolellisesti; tosin tässäkään tapauksessa ei ole kyse otannasta.

Metodologiselta kannalta yleistämiseen sisältyy aito ongelma. Tarjolla on kaksi ratkaisua: näyteaineistolla saatuja tuloksia 1) ei yleistetä vastaajajoukon ulkopuolelle, tai 2) ne yleistetään (johonkin, mihin?), mutta yleistäminen perustellaan muilla tutkimuksilla tai tutkijan muulla tietämyksellä asiasta. Ensimmäinen vaihtoehto voi kuulostaa tylsältä muttei todellakaan ole sitä. Uusi tieto kielestä on aina arvokasta. Tutkija voi pohdintaosuudessa toki tehdä yleistyksen suuntaan meneviä hypoteeseja. Joskus tieto normista ja selvistä tapauksista voi oikeuttaa yleistämisen näytteestä koko puheyhteisöön, mutta tällaisiin tapauksiin ei juuri kyselyä tarvita. Vaarallisinta on tehdä metodologisesti ei-sallittuja yleistyksiä oman kielitajun varassa tiedostamatta, mitä tekee.

Perustava ero otoksen ja näytteen välillä on se, että otoksen kohdalla tiedetään perusjoukko ja sen perusteella se, kenen piti vastata (eli ketkä valittiin otokseen), ketkä näistä vastasivat ja ketkä eivät. Näytteessä tiedetään (olettaen, että tietoa kysyttiin) vain se, ketkä vastasivat, mutta ei sitä, ketkä eivät vastanneet. Otoksessa kaikkien perusjoukkoon kuuluvien on yhtä todennäköistä päätyä mukaan tutkimukseen. Yleistämisen kannalta olennainen satunnaisuus on silloin tutkijan hallussa, ja sen mekanismit tunnetaan. Voidaan arvioida, onko kato aiheuttanut harhaa: edustavatko vastaamatta jättäneet perusjoukkoa vai muodostavatko he jonkin erityisryhmän? Jälkimmäisessä tapauksessa vastaamatta jättäminen voi olla tulosten tulkinnan kannalta olennaista.

Kielentutkijaa kiinnostaa usein se, paljonko tutkimuksessa pitäisi olla vastaajia. Ennen kuin tähän voidaan vastata, on tiedettävä, onko kyseessä satunnaisotanta vai harkinnanvarainen näyte. Satunnaisotannassa vastaajien määrä riippuu ennen muuta hyväksytystä virhemarginaalista ja ennakoidusta vastausprosentista. Näytekeruussa asiaan vaikuttavat tutkittava ilmiö ja se, halutaanko vastauksia analysoida taustamuuttujien suhteen. Tutkimuskohteesta riippuen pienikin vastaajamäärä voi hyvin riittää. Aiempien tutkimusten tarkastelu voi auttaa: Millaisia kyselyasetelmia tiedeyhteisö on aiemmin hyväksynyt julkaistavaksi? Onko tietyillä tieteellisillä julkaisuilla omia suosituksia esim. näyteaineiston koosta? Yhdysvaltojen vaaliesimerkki kertoo, että pelkkä suuri vastaajien määrä ei tee tutkimuksesta luotettavaa (ks. luku 1.1.).

Vastaajien tavoittaminen voi olla tutkimuksen pullonkaula sekä otoksessa että näytteessä. Miksi vastaaja kiinnostuisi juuri kyseisestä kielentutkimuksen kyselystä, kun verkko on täynnä erilaisia kyselyjä? Vastaamishalukkuutta voi toisinaan lisätä some-näkyvyydellä, houkuttelevalla tutkimusaiheella tai vastaajan palkitsemisella, kunhan tähän liittyvät riskit (esim. vastaajan anonymiteetin säilyminen) on otettu huomioon. Tutkija voi rahoituksen niin salliessa tilata aineistonkeruun myös alan yritykseltä. Siinä tapauksessa on otettava huomioon, että yritykset tekevät yleensä ns. kiintiöpoimintaa. Se tarkoittaa sitä, että tietyiltä alueilta kerätään vastaajakiintiöt täyteen. Kieltäytyjien määrät tai taustaprofiilit eivät ole tiedossa. Tutkija voi käyttää myös verkossa toimivia vastaamisen puoliammattilaisia (engl. *mechanical turks)*, jotka ovat sitoutuneet vastaamaan erilaisiin kyselyihin pientä korvausta vastaan. Yleistettävyyden kannalta kiintiöpoiminta tai ammattivastaajat ovat parempi kuin näyte, mutta satunnaisotantaa ne eivät vastaa.

Keskustelu verkossa ja paperilla tehtävän kyselyn hyödyistä on käytännössä ratkaistu verkon hyväksi (ks. esim. Laaksonen, Matikainen & Tikka 2013; Valli & Perkkilä 2015). Paperikysely eri vaiheineen on hidasta ja kallista. Verkkokyselyn haittatekijöitä voidaan vähentää kohdentamalla kyselyitä. Näyteaineisto voidaan kerätä verkkokyselyllä myös valvotuissa olosuhteissa esimerkiksi luennolla tai koululuokassa.

Keräämme aineiston kolmesta opetusryhmästä Helsingin yliopistolta. Ryhmämme ottaa yhteyttä luennoitsijoihin ja saa luvan käyttää keruuseen 20 minuuttia luentoaikaa. Tutkijamme menee paikalle, esittelee lyhyesti keruun ja seuraa tilannetta. Keruun jälkeen osallistujille tarjotaan tutkimuksen esittely kirjallisena. Saamme yhteensä 147 vastausta, joista 130 ilmoittaa äidinkielekseen suomen.

## 3.4. Pilottiaineiston analyysin vaikutus lomakkeen suunnitteluun

Tässä analyysia käsittelevässä luvussa nostamme esiin seikkoja, jotka ovat olennaisia lomakkeen suunnittelun kannalta. Kyselyaineiston analyysimahdollisuuksista on runsaasti kirjallisuutta, johon kannattaa tutustua jo lomaketta suunnitellessa (esim. Gómez 2009; Greenacre 2017; Gries 2013; Hand 2004; Johnson 2008; Luojola 2006; Stigler 2016).

Aineiston analysointi liittyy vahvasti jo lomakkeen suunnittelu- ja pilotointivaiheeseen. Lomakkeen jokaiseen kysymykseen liittyy suoraan suunnitelma siitä, miten odotettavissa oleva vastaus analysoidaan – ja miten analyysin tuloksella vastataan tutkimuskysymykseen. Lopullisen aineiston analyysin onnistumisen kannalta on tärkeää, että pilotointivaihe on ulotettu myös pilottiaineiston analyysiin ja analyysissa mahdollisesti ilmenneet lomakkeesta aiheutuneet ongelmat on korjattu ennen varsinaista keruuta.

Kyselymenetelmä tuottaa tyypillisesti ison määrän aineistoa, josta kielentutkijan tulee löytää tutkimuskysymyksensä kannalta olennainen tieto. Tässä auttavat tilastolliset mallit ja menetelmät. Liian ”hienoja” tai monimutkaisia tilastollisia menetelmiä kannattaa välttää etenkin, jos ei ole varma niiden sopivuudesta kyseiseen tutkimukseen. Yksinkertaiset perusmenetelmät ovat hyviä, koska niiden tulokset ovat useamman ymmärrettävissä. Kuvien piirtäminen on tärkeää, ja esimerkiksi prosentuaalisista laskelmista voi jo päätellä paljon. Laskennan pitää olla intuitiivisesti selvää ja pysyä lähellä ilmiötä. Kaikessa numeroiden yhdistämisessä tulee olla tarkkana. Liian tiivistetyt tulokset voivat hävittää niin

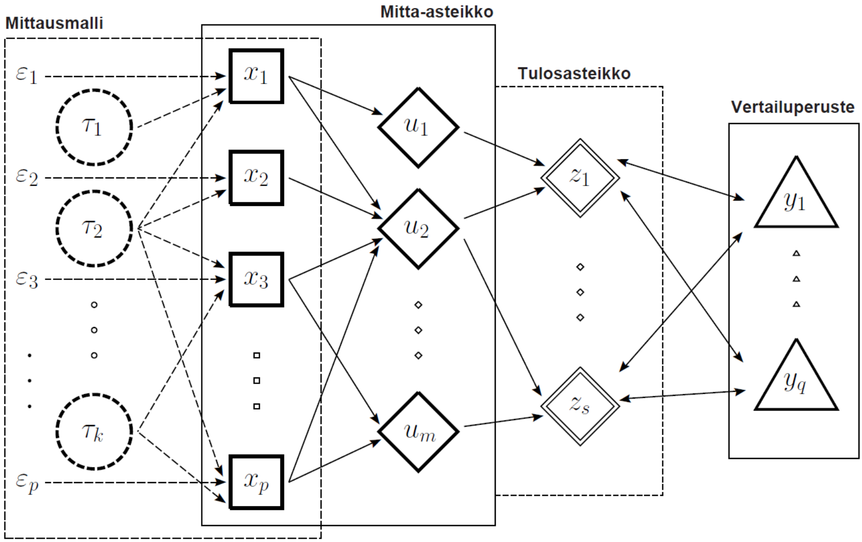
paljon yksityiskohtia, etteivät ne enää kerro ilmiöstä sitä, mitä niiden oli tarkoitus kertoa.

Kielentutkijan työpöydän yläpuolelle kiinnitetty mittausmallin tuloste (ks. kuva 1) pysyy tutkijan apuvälineenä myös analyysin suunnittelussa. Vasemman laidan ympyröiden havainnot muistuttavat siitä, mistä lähdettiin liikkeelle ja minkä perusteella luotiin tutkimuskysymykset. Jos tämä unohtuu, analyysivaiheessa tutkija voi innostua numeroista, jotka eivät välttämättä vastaakaan tutkimuskysymyksiin. Alkuperäisen mittausmallin viereen kannattaa tulostaa uusi kuva, jossa näkyy tutkimuksen kasvaminen tulosten suuntaan.

Alkuperäinen mittausmalli laajenee nyt oikeasta laidastaan mittauskehikoksi, johon tulee salmiakkikuvioita ja muita käsitteitä. Keskelle mitta-asteikoksi kehittyvät salmiakit vastaavat tyypillisesti vasemman laidan ympyröitä. Mitta-asteikon vasemmalle puolelle jää siis alkuperäinen mittausinstrumentti eli kyselylomake. Laatikot eli lomakkeen kysymykset ovat kuin peilejä, jotka kertovat jotain tutkittavasta kielen ilmiöstä. Analyysivaiheessa saatetaan huomata, että lomake toimiikin kuin Linnanmäen Naurutalon peili: se muistuttaa kuvaamaansa ilmiötä mutta yksinkertaistaa, vääristää tai muulla tavoin tulkitsee todellisuutta osin yllätyksellisesti.

Yksinkertaisimmillaan aineiston analyysi on tiivistämistä eli aggregaatiota, mihin viittaavat myös mittauskehikon asteikot, joihin tullaan suoraan alkuperäisestä aineistosta. Aggregaatio on ilmiön tuntemukseen perustuva yksinkertaistus, jota on kaikkien tilastollisten mallien ytimessä (Stigler 2016). Oletus on se, että tuloksista pitäisi löytyä paljon tietynlaisia vastauksia, jotka voidaan yhdistää ja vaikkapa esittää visuaalisesti suhteellisen selvinä kokonaisuuksina.

On huomattava sekin, että jokin kysymys voi vastata ehkä yllättäen johonkin piilevään seikkaan, vasemman laidan pieneen ympyrään, jota tutkija ei edes lähtenyt tutkimaan. On myös mahdollista, että vielä aineiston analysoinnin myötä havaitaan tarve jollekin uudelle ympyrälle eli ulottuvuudelle, jota ei tultu alun perin ajatelleeksi. Tutkijan on mahdollista työstää ajatustaan tutkittavasta ilmiöstä myös analyysin edetessä. Tällöin on kyseessä enemmän eksploratiivinen eli aineistopohjainen tutkimus kuin sen vastakohta, konfirmatorinen eli malleja ja hypoteeseja




Kuva 2. Mittauskehikko kokonaisuudessaan (Vehkalahti 2019, 122).

varmentava tutkimus. Käytännössä erilaiset tutkimusasetelmat asettuvat jonnekin näiden kahden vaihtoehdon välille.

Mittauskehikossa edetään siis vasemmalta oikealle. Useimmiten välttämättömän aineiston tiivistämisen myötä päästään lähemmäksi tutkimuskysymyksiä, kun voidaan operoida laatikoiden eli yksittäisten kysymysten sijaan ympyröitä vastaavilla salmiakeilla eli vastauksien tiivistetyillä yhdistelmillä. Yhteiskunta- ja käyttäytymistieteiden kyselyissä tyypillisin keino aineiston alkuvaiheen tiivistämiseen on faktorianalyysi (Vehkalahti 2019, 93–111), mutta vastaavaa tiivistystä voidaan tehdä yksinkertaisemminkin, esimerkiksi laskemalla muuttujia yhteen (mts. 112–116).

Tästä eteenpäin vaihtoehtoisia etenemistapoja on paljon. Analyyseihin voidaan ottaa mukaan kaavion ulkopuolisia taustatekijöitä, kuten vastaajien sukupuoli, ikä ja äidinkieli. Yhden muuttujan tarkastelut, kuten kuvien piirtäminen ja tunnuslukujen tulkitseminen, ovat aina mielekkäitä tapoja aloittaa. Seuraava taso on tarkastella kerralla kahta muuttujaa yhtaikaa. Tiivistettyjä vastauksia voidaan esimerkiksi taulukoida

ristiin taustamuuttujien kanssa ja piirtää vaikkapa pylväskuvia. Aineiston alkuvaiheen tiivistäminen on tärkeää, koska se helpottaa analysointia olennaisesti. Analysointi olisi huomattavan paljon työläämpää alkuperäisillä kysymyksillä, koska niitä on yksinkertaisesti liikaa. Mittauksia ei valitettavasti ole mahdollista suorittaa valmiiksi tiivistetysti; se johtaisi epäselvyyksiin ja väärinymmärryksiin.

Hieman pidemmälle menevistä tyypillisistä menetelmistä mainittakoon regressioanalyysi (Vehkalahti 2019, 124–150), jossa yhtä muuttujaa pyritään selittämään useilla muilla. Mittauskehikossa selittäjinä toimivat erilaiset tiivistetyt asteikot ja selitettävänä toimii jokin vertailuperusteeksi kutsuttu erillinen tekijä (kuvassa 2 ”y”). Tulokseksi saadaan muun muassa uusi tulosasteikko (regressioanalyysin tuottama ennuste, kuvassa 2 ”z”), jonka pohjalta voidaan tehdä johtopäätöksiä analyysin onnistumisesta. Analyysin jäännös- eli residuaalivaihtelu ja sen visualisointi avaavat puolestaan näkymiä siihen, kuinka hyvin regressiomalli toimii, miten sen erilaiset oletukset täyttyvät aineiston perusteella ja mitä mahdollisia lisäpäätelmiä aineiston ja mallin pohjalta voisi vielä tehdä. Vastaavia analyysimenetelmiä on lukuisia, ja ne toimivat pitkälti samoilla periaatteilla. (Perusmenetelmistä tarkemmin ks. esim. Vehkalahti 2019.)

> Esimerkkitutkimuksessamme oikeiden ja väärien vastausten määrien tarkastelu antaa ns. ensimmäisen asteen tuloksia. Oikeiden vastausten lisäksi aineistossa on paljon vääriä vastauksia, joissa on lukuisia tapoja ”ei-tietämiselle”. Jäännösvaihtelusta saattaa löytyä uusia näkökulmia tutkittavaan asiaan, ja sitä kannattaa tarkastella ihan uteliaisuudesta. Näitä kutsutaan myös toisen asteen tuloksiksi. Yksi tapa analysoida aineistoa on myös tarkastella tuloksia vastaajittain. Montako oikeaa ja väärää vastausta kukin vastaaja tuotti? Onko joukossa vastaajia, jotka vastasivat kaikki oikein? Löytyykö systemaattisesti oikein tai väärin vastanneille yhteisiä taustatekijöitä, ja voiko tästä päätellä jotakin?
>
> Olemme lisänneet lomakkeeseemme kysymyksiä ja redundanssia. Samalla satunnaisvaihtelun osuus tuloksissa on kasvanut. ”Jokaisen sanan kohdalle on tulossa valtavan harva matriisi ykkösiä

ja nollia. Jotenkin tuota pitäisi tiivistää ja runnoa kasaan, että sieltä nähtäis jotain", sanoo tilastotieteilijä.

Aluksi laskemme binaarisesti, montako prosenttia vastaajista on vastannut mihinkin kysymykseen oikein ja montako väärin. Tällä yksinkertaisella erottelulla saamme ns. ensimmäisen asteen tulokset ja vastauksen tutkimuskysymykseemme: tunnistavatko vastaajat tunnesanojen alkusanakielisiä määritelmiä? Toisen asteen tuloksia saamme jäännösvaihtelusta. Luomme tiedoston, josta poistamme oikeat vastaukset. Saamme vielä edellistä harvemman matriisin, josta voimme tarkastella sitä, mikä on vaikkapa kolmen kärki väärissä vastauksissa per määritelmä (olettaen, että enemmistö vastaajista vastasi oikein). Näemme myös, mitä sanoja on valittu vähiten.

Teemme jokaiselle seitsemälle tunnesanalle nämä kaksi toimenpidettä. Jos voisimme luottaa siihen, että mittaus on täysin virheetön, yllä olevat tulokset riittäisivät meille. Olemme kuitenkin varautuneet mittausvirheisiin kasvattamalla lomakkeen redundanssia osakomponenteista muodostetuilla lisäkysymyksillä. Osakomponentteja emme voi tarkastella aivan samalla tavalla kuin kokonaismääritelmiä, koska yhtä osakomponenttia voi vastata useampi tunnesana. Tähän tarvitsemme vielä lisätarkasteluja aineiston äärellä.

Teemme myös osakomponenteille primaarianalyysin eli katsomme kysymys kerrallaan, moniko vastaaja on vastannut oikein. Tuloksia voimme myöhemmin verrata kokonaismääritelmien tunnistusprosentteihin. Onko kokonaismääritelmä mahdollisesti paremmin tai heikommin tunnistettu kuin yksittäiset merkityskomponentit? Puuttuuko kyselyn pohjalla olevasta nojatuolimääritelmästä jotain olennaista?

On tärkeää, että tilastotieteilijän kanssa käydään keskustelua jo lomakkeen suunnitteluvaiheessa. Harmillisen usein tilastotieteilijä kutsutaan apuun liian myöhään, jolloin iso osa kielentutkijan tekemästä työstä voi joutua hukkaan. Myös opiskelijan on selvitettävä, kenen puoleen voi tilastoasioissa kääntyä. Lisäksi on muistettava arkijärjen käyttö. Hälytyskellojen pitäisi soida, jos kielentutkija ei lainkaan ymmärrä, mitä

tilastotieteilijä tekee tai käskee tehdä kieliaineistolla. Keskustelut tilastotieteilijän kanssa avaavat näkökulmia, antavat työkaluja ja varoittavat vaarasta, mutta vastuu prosessin ymmärtämisestä ja tulosten tulkinnasta jää aina kielentutkijalle.

Keskustelu tilastotieteilijän kanssa voi aiheuttaa myös ristiriitoja. Kielentutkija voi osata käyttää Exceliä, mutta tilastotieteilijän mukaan sillä voi olla vaikeaa tehdä tilastollista tutkimusta, koska se ei ole siihen varsinaisesti tarkoitettu. Kielentutkijan pitäisi käyttää muutama viikko tai kuukausi esimerkiksi R:n opiskeluun[2]. Tavoitteista riippuen tämä kannattaa tai ei kannata. Exceliä kehittyneemmän työkalun opettelu voi kielentutkijoiden kokemusten mukaan olla alkuun hidasta, mutta kun systeemin oppii, käyttö voi olla todella nopeaa. Vaikka verkosta löytyykin kosolti hyviä neuvoja, voi aikaa vievintä silti olla erilaisten komentojen etsiminen tutoriaaleista tai sen keksiminen, mikä on se nappi, jota painamalla kaikki (muka!) järjestyy kuten halutaan.

## 3.5. Tulosten tulkinta ja kirjoittaminen

Tulosten tulkinnan ja kirjoittamisen vaiheessa nähdään lopulta se, onnistuivatko kysymykset mittaamaan sitä, mitä alussa lähdettiin tutkimaan. Ovatko tulokset odotusten mukaisia vai yllättävätkö ne? Painottuvatko tulokset jotenkin? Onko jokin mittausmallin vasemman laidan ympyröistä tulosten perusteella kiinnostavampi kuin alussa luultiin? Lisäksi on muistettava kiinnittää kriittistä huomiota aineiston keruun kuvaukseen ja tulosten yleistettävyyteen.

Kyselytutkimuksen raportoinnin ongelmat liittyvät usein aineiston keruun kuvaukseen (Laaksonen 2017). Lukijalle jää toisinaan epäselväksi jopa se, onko kyseessä ollut otanta vai harkinnanvarainen näyte. Otannan ja näytteen eron lyhyt selvittäminen auttaisi myös lukijaa tulkitsemaan tuloksia. Ongelmia voi liittyä myös aivan asianmukaisesti kerätyn otoksen kuvaukseen: otoksen suuruutta voidaan korostaa, vaikka pelkkä vastaajien lukumäärä ei kerro paljoakaan, jos vastaajien suhteellinen määrä eli vastausprosentti jää ilmoittamatta. Ongelmista vaikeneminen ei aina ole tahallista, vaan kyse voi olla tiedon ja osaamisen puutteista.

Ongelmia ei välttämättä ole edes tunnistettu. Jos otantaan liittyvät ongelmat tunnistettaisiin, osa näistä voitaisiin analyysivaiheessa oikaista. Esimerkiksi kadon aiheuttamia ongelmia voitaisiin korjata painotuksilla tai kuvaamalla vastanneiden ja vastaamatta jättäneiden profiileja. Valmiiden aineistojen varaan rakentuvassa tutkimuksessa on otettava huomioon mittaamiseen aikanaan mahdollisesti liittyneet ongelmat.

Yleistämiseen puolestaan liittyy usein se ongelma, että harkinnanvaraisella näytteellä kerätyn aineiston tulokset yleistetään koskemaan jotakin perusjoukkoa, vaikka ne edustavat vain sitä vastaajaryhmää, josta aineisto on kerätty. Yleistäminen ei ole ilmoitusasia, eikä se toisaalta aina ole edes tarpeen. Huolellisesti raportoidut tulokset ovat arvokkaita jo sellaisenaan: on saatu selvitettyä ilmiön perusasiat ja testattua laadittu kyselylomake käytännössä. Tiedetään, miten tulokset poikkesivat tutkijan ennakkokäsityksistä. Jos näyteaineiston tuloksia yleistetään näytteen ulkopuolelle, yleistäminen nojaa tällöin tutkijan muuhun tietämykseen ja mahdollisiin muilla menetelmillä kerättyihin aineistoihin. On myös tärkeää muistaa, että kyselyssä vastaajalta saatu tieto ei välttämättä vastaa sitä, mitä vastaaja tuottaisi spontaanissa kielenkäyttötilanteessa.

Mitään ei pystytä mittaamaan täysin virheettömästi. Tämä ei kuitenkaan estä uuden tiedon saavuttamista. Eettisesti toimiva kielentutkija on tietoinen kyselyynsä liittyvistä ongelmista, pyrkii niiden totuudenmukaiseen raportointiin, osoittaa tulostensa yhteydet aiempaan tutkimukseen ja käy näin keskustelua tiedeyhteisön ja yhteiskunnan kanssa.

Joskus käy niin, että kaikista varotoimenpiteistä huolimatta tutkija jollain tavoin epäonnistuu kyselyasetelmansa kanssa. Tuloksia analysoitaessa ja kirjatessa voidaan yrittää pelastaa se, mitä pelastettavissa on. Alkuperäiseen tutkimuskysymykseen ei ehkä saada vastausta, mutta aineisto voi antaa vastauksia johonkin muuhun. Etenkin opinnäytetöiden ollessa kyseessä voidaan miettiä luovasti, miten kandidaatti tai maisteri valmistuisi määräajassa. Suuren luokan katastrofien pitäisi lähtökohtaisesti olla vältettävissä jo opinnäytetöistä alkaen.

> Tuloksia tulkitessa on aika tarkastella kriittisesti myös kyselylomakettamme. Pohdimme, mikä merkitys on sillä, että annoimme vastausvaihtoehdoiksi vain julkaistuihin kokonaismääritelmiin

sisältyneet komponentit, emme muita (saati kaikkia) mahdollisia komponentteja. Mietimme myös täytesanojen valinnan onnistumista. Huomaamme, että vaihtoehdoissa aakkosten alkupää (*ahdistus*) näyttää saavan enemmän rasteja kuin aakkosten loppupään vaihtoehdot. Vastaajat eivät ehkä ole jaksaneet käydä läpi kaikkia vaihtoehtoja. Toteamme, että tätä vääristymää voisi ehkä yrittää oikaista tilastotieteen keinoin, mutta välttämättä se ei onnistu. Löydämme edelleen hyvät perustelut kaikille lomakkeemme isoille ratkaisuille, mutta joitakin yksittäisiä sanavalintoja voisimme tehdä nyt toisin. Toteamme samalla, että tunnesanojen alkusanakielisissä selityksissä riittää tutkittavaa myös jatkossa ja myös toisen tyyppisillä lomakkeilla.

## 4. Yhteenveto

Kuten lukija on ehkä huomannut, otteemme kyselymenetelmään on paikoitellen melko normatiivinen. Toivomme tämän auttavan menetelmän omaksumisessa. Emme kuitenkaan ole halunneet karkottaa herkimpiä lukijoita ylenmääräisellä ankaruudella, vaan tahtomme on ollut ystävällisesti ohjata tulevia tutkijoita pahimpien sudenkuoppien ohi. Kyselyn tekeminen on parhaimmillaan hulvattoman hauskaa, ja ensimmäisten tulosten ilmaantuminen tietokoneen ruudulle on jännittävämpää kuin lahjojen avaaminen jouluaattona!

Kysely on kielentutkijalle erinomainen tutkimusmetodi, jolla on paljon etuja. Sen avulla voidaan kielestä tai kieleen liittyvistä käsityksistä saada sellaista uutta tietoa, jota muilla menetelmillä on hankalaa tai jopa mahdotonta saada. Kyselyllä voidaan kerätä laaja ja juuri kyseessä olevaa tutkimuskysymystä varten suunniteltu aineisto. Kyselyä voidaan sujuvasti yhdistellä myös muihin menetelmiin.

Kyselyn käyttöön liittyviä riskejä pystytään huolellisella suunnittelulla ennakoimaan. Tärkeintä on huomata, että näennäisestä yksinkertaisuudestaan huolimatta kysely ei ole menetelmänä ihan helppo.

Onnistuneen kyselyn toteuttamiseen tarvitaan riittävästi aikaa ja viitseliäisyyttä. Kyselyn menestyksekäs käyttö vaatii kielitieteellisen osaamisen lisäksi vähintään jonkinasteista tietoa tilastotieteestä, ehkä myös psykologiasta. Tietoa on kuitenkin runsaasti ja helposti saatavilla menetelmäoppaissa.

Ihannetapauksessa kyselyä harjoitetaan ryhmässä, jossa on monipuolista osaamista ja jossa tehtäviä voidaan jakaa. Opinnäytetyötä tekevän on – etenkin tukijoukkojen puuttuessa – suositeltavaa tukeutua aiemmin käytettyihin kyselyasetelmiin. Järkevä vaihtoehto voisi olla myös kerätä uusi aineisto opiskelijaryhmässä, jossa kukin analysoi osan tuloksista ja kirjoittaa opinnäytteensä osa-aineiston perusteella. Tai voisiko graduryhmän joskus perustaa menetelmän eikä kielenilmiön tai oppiaineen ympärille?

Kyselytutkimuksen teoria ja ideaalit ovat omat todellisuutensa, ja käytännössä on etsittävä tapa tehdä tutkimusasetelma järkevästi. On tehtävä kompromisseja ja luovuttava siitä hienoimmasta mahdollista lomakkeesta, jota yksikään vastaaja ei jaksaisi täyttää asiallisesti loppuun. Sama pätee aineiston analyysissa. Parhaimmillaan kyselytutkimus tuottaa kiinnostavaa ja luotettavaa tietoa suoraan vastaajilta tutkijalle. Tulokset valottavat tutkijan ennalta keksimiä kysymyksiä ja tuovat tutkittavaan ilmiöön uusia, yllättäviäkin näkökulmia.

**Aiheesta lisää:**

**Kyselytutkimuksen perusteista ja käytännöistä:**

Croves, Robert M., Fowler, Floyd J. Jr., Couper, Mick P., Lepkowski, James M., Singer, Eleanor, Tourangeau, Roger. 2009. *Survey Methodology.* (2. p.). Hoboken: John Wiley & Sons, Inc.

Laaksonen, Seppo. 2013. *Surveymetodiikka. Aineiston kokoamisesta puhdistamisen kautta analyysiin.* Kööpenhamina: Ventus Publishing. Saatavissa: http://bookboon.com/fi/surveymetodiikka-ebook.

Vehkalahti, Kimmo. 2019. *Kyselytutkimuksen mittarit ja menetelmät*. Helsinki: Helsingin yliopisto. Saatavissa: http://doi.org/10.31885/9789515149817.

**Kyselyaineiston keruusta kielentutkimuksessa:**

Buchstaller, Isabelle & Khattab, Ghada. 2013. Data sampling in linguistics. Julkaisussa: Podesva, Robert J. & Sharma, Jevyani (toim.) *Research Methods in Linguistics*. Cambridge: Cambridge University Press, 74–95.

## VIITTEET

1 Tapaustutkimukseemme voi tutustua tarkemmin artikkelissa Vanhatalo ja muut (2019).
2 https://www.google.com/search?q=r+for+beginners

## KIRJALLISUUS


Ainiala, Terhi & Lappalainen, Hanna. 2010. Miten Helsingissä puhutaan. *Virittäjä* 114:1, 71–107.

Alastalo, Marja. 2005. *Metodisuhtanteiden mahti. Lomaketutkimus suomalaisessa sosiologiassa* 1947–2000. Tampere: Vastapaino.

Alkula, Tapani, Pöntinen, Seppo & Ylöstalo, Pekka. 1994. *Sosiaalitutkimuksen kvantitatiiviset menetelmät*. Porvoo: WSOY.

Buchstaller, Isabelle & Khattab, Ghada. 2013. Data sampling in linguistics. Julkaisussa: Podesva, Robert J. & Sharma, Jevyani (toim.) *Research Methods in Linguistics*. Cambridge: Cambridge University Press, 74–95.

Chambers J. K. & Trudgill, Peter. 1980. *Dialectology*. Cambridge: Cambridge University Press.

Croves, Robert M., Fowler, Floyd J. Jr., Couper, Mick P., Lepkowski, James M., Singer, Eleanor, Tourangeau, Roger. 2009. *Survey Methodology*. (2. p.). Hoboken: John Wiley & Sons, Inc.

Foddy, William. 2001. *Constructing Questions for Interviews and Questionnaires. Theory and Practice in Social Research*. (6. p.). Cambridge: Cambridge University Press.

Fowler, Floyd J. Jr. 1995. *Improving Survey Questions: Design and Evaluation*. London: Sage.

Goddard, Cliff & Wierzbicka, Anna. 2014. *Words and Meanings: Lexical Semantics Across Domains, Languages, and Cultures*. Oxford: Oxford University Press.

Gómez, Pascual Cantos. 2009. *Statistical Methods in Language and Linguistic Research*. Oakville (CT): Equinox Pub. Ltd.

Greenacre, Michael. 2017. *Correspondence Analysis in Practice*. (3. p.). Boca Raton: CRC Press.

Gries, Stefan Th. 2013. *Statistics for Linguistics with R: A Practical Introduction*. (2. p.). Berlin: Mouton de Gruyter.

Hand, David J. 2004. *Measurement Theory and Practice: The World Through Quantification*. London: Hodder Arnold.
Hirsjärvi, Sirkka, Remes, Pirkko & Sajavaara, Paula. 2009. *Tutki ja kirjoita*. (15. uud. p.). Helsinki: Tammi.
Itkonen, Esa & Pajunen, Anneli. 2010. *Empiirisen kielitieteen metodologia*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Johnson, Keith. 2008. *Quantitative Methods in Linguistics*. Oxford: Blackwell.
Jürvetson, Maria-Magdalena. 2015. *Suomen kielen nauramisverbit koloratiivikonstruktiossa: muoto, merkitys ja tehtävät*. Väitöskirja. Tallinna: Tallinnan yliopisto.
Laaksonen, Salla-Maaria, Matikainen, Janne & Tikka, Minttu (toim.). 2013. *Otteita verkosta*. Tampere: Vastapaino.
Laaksonen, Seppo. 2013. *Surveymetodiikka. Aineiston kokoamisesta puhdistamisen kautta analyysiin*. Kööpenhamina: Ventus Publishing. Saatavissa: http://bookboon.com/fi/surveymetodiikka-ebook.
Laaksonen, Seppo. 2017. Tutkijan hallittava otostutkimuksen perusteet. *Tieteessä tapahtuu*. 35:4, 55–56.
Labov, William, Ash, Sharon, & Boberg, Charles. 2006. *Atlas of North American English: Phonetics, Phonology, and Sound Change*. New York (NY): Mouton de Gruyter.
Leino, Pirkko. 1968. Suomalaiset matkimukset. Pro gradu -tutkielma. Helsingin yliopisto, suomen kieli.
Leppänen, Sirpa, Pitkänen-Huhta, Anne, Nikula, Tarja, Kytölä, Samu, Törmäkangas, Timo, Nissinen, Kari, Kääntä, Leila, Räisänen, Tiina, Laitinen, Mikko-Pekka, Pahta, Päivi, Koskela, Heidi, Lähdesmäki, Salla & Jousmäki, Henna. 2011. *National Survey on the English Language in Finland: Uses, Meanings and Attitudes*. Studies in Variation, Contacts and Change in English, vol 5. Helsinki: Research Unit for Variation, Contacts and Change in English.
Leskinen, Heikki. 1991. Vieläkö nuoret nurisevat? Huomioita onomatopoeettisten sanojen tuntemuksesta ja tulkinnasta. *Virittäjä* 95:4, 355–371.
Luodonpää-Manni, Milla. 2013. Kyselytutkimus kielentutkijoiden yleistieteellisille termeille antamista merkityksistä. *Virittäjä* 117:2, 242–262.
Luojola, Teemu. 2006. *Kielitieteellisen aineiston kvantitatiiviset analyysimenetelmät*. Moniste, Helsingin yliopisto, yleinen kielitiede. Saatavissa: http://www.ling.helsinki.fi/~fkarlsso/methods/kvant_men.pdf.
Mielikäinen, Aila & Palander, Marjatta. 2002. Suomalaisten murreasenteista. *Sananjalka* 44, 86–109.
Priiki, Katri. 2011. Läsnäolo, tuttuus, tulkinta ja sävy. Ihmisviitteinen *hän* ja sen tehtävät Harjavallan, Kokemäen ja Huittisten nykypuhekielessä. Pro gradu -tutkielma. Turun yliopisto, suomen kieli.
Raukko, Jarno. 2002. Pitämisen polysemia: Miten koehenkilöt hahmottavat *pitää*-verbin merkitystyyppejä. *Virittäjä* 118:3, 354–374.
Räsänen, Maaret & Palander, Marjatta. 2015. Kansandialektologinen testi murrepiirteiden keskinäisestä murteellisuusjärjestyksestä. *Virittäjä* 119:1, 4–33.
Saris, Willem E. & Gallhofer, Irmtraud N. 2007. *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. Hoboken: Wiley.

Schilling, Natalie. 2013. Surveys and interviews. Julkaisussa: Podesva, Robert J. & Sharma, Jevyani (toim.) *Research Methods in Linguistics.* Cambridge: Cambridge University Press, 96–115.

Siitonen, Kirsi. 2016. Ihmistarkoitteinen *hän*-pronomini puhutussa suomessa – kansanlingvistinen kyselytutkimus. *Acta Philologica* 49, 383–392.

Spearman, Charles. 1904. General intelligence, objectively determined and measured. *American Journal of Psychology* 15, 201–293.

Stigler, Stephen M. 2016. *The Seven Pillars of Statistical Wisdom*. Cambridge: Harvard University Press.

Thurstone, L. L. 1931. Multiple factor analysis. *Psychological Review* 39, 406–427.

Tummers, Jose, Heylen, Kris & Geeraerts, Dirk. 2005. Usage-based approaches in Cognitive Linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory* 1:2, 225–61.

Valli, Raine & Perkkilä, Päivi. 2015. Nettikyselyt ja sosiaalinen media aineistonkeruussa. Julkaisussa: Aaltola, Juhani & Valli, Raine (toim.) *Ikkunoita tutkimusmetodeihin I – metodin valinta ja aineiston keruu: virikkeitä aloittelevalle tutkijalle.* (4., uud. p.). Jyväskylä: PS-Kustannus, 109–120.

Vanhatalo, Ulla. 2002. ”Naiset motkottaa aiheesta ja nalkuttaa syyttä”. Kyselytesti verbien semanttisten sisältöjen arvioinnissa. *Virittäjä* 106:3, 330–353.

Vanhatalo, Ulla. 2005. *Kyselytestit synonymian selvittämisessä. Sanastotietoutta kielenpuhujilta sähköiseen sanakirjaan*. Helsinki: Helsingin yliopisto.

Vanhatalo, Ulla & Tissari, Heli. 2017. Esittelyssä alkusanakieli. *Virittäjä* 121:2, 244–263.

Vanhatalo, Ulla, Vehkalahti, Kimmo, Lilja, Taru, Siiroinen, Mari & Tissari, Heli. 2019. ’Something bad can happen to me here now’: Meaning components of emotion words. *SKY Journal of Linguistics* 32, 145–179.

Vastamäki, Jaana. 2015. Kyselylomaketutkimus: Tutkimusasetelman ja mittareiden valinta. Julkaisussa: Aaltola, Juhani & Valli, Raine (toim.) *Ikkunoita tutkimusmetodeihin I – metodin valinta ja aineiston keruu: virikkeitä aloittelevalle tutkijalle.* (4., uud. p.). Jyväskylä: PS-Kustannus, 121–132.

Vehkalahti, Kimmo. 2019. *Kyselytutkimuksen mittarit ja menetelmät*. Helsinki: Helsingin yliopisto. Saatavissa: http://doi.org/10.31885/9789515149817.

Västi, Katja. 2012. Elävä LÄHDE. Alkuasemaisen ablatiivin merkitystyypit verbittömässä konstruktiossa. *Virittäjä* 116:1, 67–97.

# Luku 3 Haastattelu aineistonkeruun metodina sosiolingvistiikassa

*Tommi Kurki*
https://orcid.org/0000-0003-2858-9956
*Liisa Mustanoja*
https://orcid.org/0000-0003-1818-6926

**Mikä?**

*Sosiolingvistinen haastattelu* tarkoittaa väljästi tulkittuna mitä tahansa keruutapaa, jossa tallennetaan puheaineistoa kasvokkaistilanteessa sosiolingvistisen tutkimuksen tarpeisiin. Tarkoituksena on yleensä saada analysoitavaksi informanteilta runsaasti ns. vapaamuotoista puhetta. Toisinaan käsitteellä viitataan vielä nykyäänkin William Labovin kehittämään keruuproseduuriin, jossa puhetta pyritään keräämään useasta eri kontekstuaalisesta tyylistä, toisinaan tästä edelleen kehiteltyyn keruutapaan tai esimerkiksi tapaan, jossa puheen annetaan virrata hyvin vapaasti.

**Katso myös:**

Luku 2 Kysely

Luku 5 Lingvistinen etnografia toimintayhteisöissä

Luku 9 Korpusaineistot

Luku 13 Kielten dokumentointi ja kieliopin kuvaus

Luku 21 Sosiolingvistiikka

# 1. Johdanto

Sosiolingvistisessä variaationtutkimuksessa ja ylipäätään sosiolingvistiikassa käytetään useita aineistonkeruumenetelmiä, esimerkiksi haastatteluja, kirjallisia kyselyjä, kenttätutkimuksen yhteydessä tehtäviä kyselyjä, nopeita anonyymejä keruita ja osallistuvaa havainnointia (ks. esim. Milroy & Gordon 2003, 52–71; Sosiolingvistiikka, Lingvistinen etnografia toimintayhteisöissä, Kielten dokumentointi ja kieliopin kuvaus tässä kirjassa). Se, mikä on hyvää aineistoa ja miten sitä pitäisi hankkia, riippuu täysin tutkimuksen tavoitteista ja tutkimuskysymyksistä (Milroy & Gordon 2003, 49).

Edellä mainituista menetelmistä haastattelu on perinteinen tapa kerätä puheaineistoa sosiolingvistiikan lisäksi myös dialektologiassa. Sitä sovelletaan myös useilla muilla kielentutkimuksen aloilla ja yleisemminkin eri tieteenaloilla, erityisesti ihmistieteissä. Haastattelun on arveltu olevan jopa käytetyin tiedonhankkimisen menetelmä yhteiskuntaelämän eri alueilla (Tiittula & Ruusuvuori 2009, 9).

Vaikka haastattelu menetelmänä ja nimityksenä on yleinen, ovat tutkijat käyttäneet metodinaan tarkemmin tarkasteltuna monenlaisia haastatteluja ja haastattelun merkitys on erilainen eri yhteyksissä. Haastattelujen eri lajeja on mahdollista tyypitellä ensinnäkin sen mukaan, kuinka tarkkaan ennen haastattelua on päätetty ja muotoiltu kysymykset ja mahdolliset vastausvaihtoehdot, mahdollisesti myös missä järjestyksessä kysymykset esitetään sekä kuinka sitovasti suunniteltuja ratkaisuja noudatetaan. Tämän perusteella haastattelut on tapana jakaa karkeasti strukturoituihin, puolistrukturoituihin ja strukturoimattomiin haastatteluihin (Tiittula & Ruusuvuori 2009, 11)[1].

Nimensä mukaisesti strukturoitu haastattelu on näistä sidotuin haastattelulaji, jossa tiukimmillaan lomakemuotoisessa haastattelussa on tarkkaan muotoillut kysymykset ja vastausvaihtoehdot sekä kysymysten järjestys on vakioitu, jotta tutkija vaikuttaisi mahdollisimman vähän haastatellun vastauksiin. Strukturoitu haastattelu voidaan toteuttaa esimerkiksi kokonaan suullisesti niin, että haastattelija esittää kysymyksensä suunnitellusti, haastateltava vastaa niihin ja haastattelu tallennetaan.

Vaihtoehtoisesti strukturoitu haastattelu voidaan toteuttaa niin, että haastattelija kysymysten esittämisen lisäksi myös kirjaa haastateltavan vastaukset.

Strukturoimaton haastattelu puolestaan edustaa haastattelulajien jatkumon toista ääripäätä, ja se muistuttaa eniten vapaamuotoista keskustelua. Haastattelu ei etene niinkään kysymys–vastaus-parien mukaisesti, vaan sekä haastateltava että haastattelija voivat kuljettaa sitä toivomaansa suuntaan ja nostaa mieleensä tulevia asioita esiin. Tarkkojen kysymysten esittämisen sijasta pyritään keskustelunomaisuuteen ja luonnollisuuteen, jotta haastatellulla olisi mahdollisuus kertoa kokemuksistaan, tuntemuksistaan, muistoistaan, mielipiteistään ja perusteluistaan (Hirsjärvi & Hurme 2000; Saaranen-Kauppinen & Puusniekka 2006). Sanatarkasti strukturoimaton haastattelu ei ole kuitenkaan koskaan täysin strukturoimaton, koska haastattelulla on aina jokin tarkoitus tai päämäärä ja usein esimerkiksi sen teema tai haastattelutilanne on etukäteen jotenkin sovittu tai määritelty.

Puolistrukturoiduissa haastatteluissa, jotka asettuvat edellä esitettyjen ääripäiden väliin, on etukäteen päätetty ja lyöty lukkoon vain osa niistä kaikista tekijöistä, jotka on vakioitu strukturoidussa haastattelussa. Yksi yleisimmistä tämän haastattelulajin edustajista on teemahaastattelu, jossa haastattelua varten keskusteltaviksi suunnitellut aihepiirit ovat samoja kaikille haastateltaville (ks. Hirsjärvi & Hurme 2000, 47–48, 66; Saaranen-Kauppinen & Puusniekka 2006).

Haastatteluja on tapana ja mahdollista luonnehtia myös sen mukaan, onko haastateltavia vain yksi, kaksi vai useita. Tämän perusteella käytetään sellaisia nimityksiä kuin *yksilöhaastattelu*, *parihaastattelu* ja *ryhmähaastattelu*. Jako voi kertoa myös tarkemmasta haastattelun tekemisen tekniikasta: virikehaastattelussa mukana on esimerkiksi kuvamateriaalia, puhelinhaastattelussa vuorovaikutustilanne ei ole kasvokkainen, kävelyhaastattelussa taas kuljetaan rinta rinnan samalla keskustellen (ks. esim. Törrönen 2017; Bamberg 2017; Ikonen 2017). Yhtä lailla haastatteluja voidaan luokitella sen perusteella, millaiseen tietoon haastattelu kohdistuu. Esimerkiksi narratiivisissa eli kerronnallisissa haastatteluissa haastateltava kertoo kokemuksistaan; etnografinen haastattelu kohdentuu haastateltavan elämänpiiriin. (Tiittula & Ruusuvuori 2009, 12.)

Kuten edellä esitetystä on käynyt jo ilmi, käytetään erilaisia haastatteluja hyvin monella eri tieteen- ja tutkimusalalla. Tässä artikkelissa ei ole kuitenkaan mahdollista käsitellä menetelmää sen kaikessa kirjossaan, ja keskitymmekin siksi seuraavassa siihen, miten sitä käytetään sosiolingvistiikassa.[2] Haastattelua on hyödynnetty dialektologiassa ja sosiolingvistiikassa ennen muuta, jotta tarkasteltavaksi on saatu kerättyä runsaasti puheyhteisön jäsenten puhetta. Esimerkiksi perinteisissä murrekeruissa kielentutkijaa eivät ole ensisijaisesti kiinnostaneet informanttien tiedot paikallisesta kansanperinteestä vaan se, kuinka he puhuvat. Tosin murteentutkimuksessakin informantin taustaa koskevat tiedot ovat olleet tärkeitä tutkijalle. Viimeistään sosiolingvistisen asennetutkimuksen ja erityisesti kansanlingvistisen tutkimuksen myötä haastatteluaineistossa on alkanut yhä enemmän kiinnostaa sen ohella, miten haastateltava puhuu, myös se, mitä hän puhuessaan kertoo. Sen sijaan esimerkiksi kasvatustieteilijöitä tai sosiologeja ei tyypillisesti kiinnosta haastatteluaineistossa se, miten haastateltavat kieltä käyttävät, vaan niiden asiasisältö. Tässä artikkelissa tarkastelemme haastattelua ennen muuta puhutun kielen tutkijoiden näkökulmasta. Laveammin ja kattavammin haastattelua käsitellään muissa teoksissa, joista usea keskittyy tarkastelemaan yksinomaan haastattelua (ks. esim. Hirsjärvi & Hurme 2000; Ruusuvuori & Tiittula 2009; Ruusuvuori, Nikander & Hyvärinen 2010; Hyvärinen, Nikander & Ruusuvuori 2017).

# 2. Haastattelumetodin historiaa

## 2.1. Dialektologia ja haastattelu

Puhutun kielen tutkimuksessa haastattelut ovat edustaneet aineistonkeruun valtavirtaa jo perinteisessä murteentutkimuksessa[3], jota on harjoitettu tieteellisen kielentutkimuksen alkuajoista 1800-luvun lopulta (ks. esim. Juusela 1994, 1–2; Chambers & Trudgill 1998; Kurki & Mustanoja 2019). Kielitieteissä haastattelua on käytetty menetelmänä myös muun muassa sanasto- ja nimistökeruissa (ks. Nimistöntutkimus tk.)

Haastattelu kytkeytyi perinteisessä tutkimuksessa alkuaikoina yleensä haastatteluissa käytettyihin sanaluetteloihin ja kyselyihin, jotka usein olivat elisitoituja (ks. esim. Chambers & Trudgill 1998, 15–23; kyselyistä yleisemmin ks. Kysely tk.; elisitoinnista Kielten dokumentointi ja kieliopin kuvaus tk.).

Myös Suomessa haastattelu on ollut murteentutkimuksessa väljästi määritellen yksi keskeisimmistä keruumenetelmistä 1800-luvun lopulta lähtien. Varhaisimmissa keruissa käytettiin keräilysanastoja, ja ensimmäiset ohjeistetut keruumenetelmät koskivat nimenomaan näiden sanaluettelojen hyödyntämistä. Toisaalta Suomessa tehtiin huomioita myös yleisemmin haastateltavien puheesta, ja silloista täkäläistä menetelmää on luonnehdittu joskus kysele ja kuuntele -metodiksi. (Ks. Strandberg 2004, 27–29.)

Nauhurit yleistyivät murteentutkijoiden käytössä 1950-luvulla (Chambers & Trudgill 1998, 24–25), minkä myötä haastattelun merkitys vahvistui entisestään. Jo tätä ennen oli murteen tallentamisessa käytetty jonkin verran muita laitteita 1900-luvun alkupuolella, mutta niitä tutkijan oli verrattain vaikeaa kuljettaa mukanaan keruumatkoillaan. Siksi murteen tallentaja joutui 1950-luvulle asti tukeutumaan haastattelussa tekemiinsä muistiinpanoihin. (Ks. esim. Yli-Paavola 1970, 9–11.) Nauhureiden yleistymisen myötä tavoitteeksi tuli tallentaa laajemmin mahdollisimman vapaamuotoista puhetta. Analyysissakaan ei oltu enää yksittäisen tutkijan ainutkertaisen kuulovaikutelman varassa (Chambers & Trudgill 1998, 24–25; nykyisistä tallennusvälineistä ja -tavoista ks. Kielten dokumentointi ja kieliopin kuvaus; Kokeellinen fonetiikka tk.). Yleisesti ottaen ero varhemmilla 1900-luvun alun välineillä tehtyihin verrattain harvinaisiin tallenteisiin oli huomattava, koska näitä ensimmäisiä kielitieteellisiä näytteitä on luonnehdittu joskus etukäteen valmisteltujen esitysten kaltaisiksi, ”esitysnäytteiden luonteisiksi” (Yli-Paavola 1970, 2, 10–12, 43).

Suomessa voi katsoa järjestelmällisen suomen murteiden äänitystyön aloitetun viimeistään vuonna 1959, jolloin perustettiin SKNA eli Suomen kielen nauhoitearkisto (ks. tietolaatikko 1). Vaikka muun muassa Suomessa huomio oli kohdistunut jo ennen nauhureita toki muihinkin kuin yksittäisiin sanoihin[4], tarjosivat nauhoitteet tutkijoille uudenlaista

monipuolista aineistoa muun muassa lause- ja tyyliopin tarkasteluun ja keinoja myös puheen painotuksen, intonaation, puherytmin, äänteiden keston ja laadun kokeellisfoneettiseen tutkimukseen. Osaltaan ne sisälsivät luonnollisissa käyttöyhteyksissään runsaasti myös muoto-opillista, sanastollista ja fraseologista aineistoa (Yli-Paavola 1970, 10–12).

## 2.2. Klassinen sosiolingvistinen haastattelu

Sosiolingvistinen haastattelu muistuttaa monin tavoin edellä esitettyä murteentutkimuksessa käytettyä haastattelua. Niissä maissa, joissa murreaineistoa kerättiin lähinnä kyselyin ja sanaluetteloin tai kysymällä keruuretkillä suoraan tietoja yksittäisistä sanoista tai rakenteista, siirtymä vanhoista menetelmistä vapaamuotoisen puheen tarkastelemiseen haastattelun keinoin oli huomattava (ks. Kielten dokumentointi ja kieliopin kuvaus tk.). Sen sijaan esimerkiksi puhesuomen tutkimuksessa siirtymä sosiolingvistiseen haastatteluun oli vähäisempi (ks. esim. Mielikäinen 1982, 278–279), koska fennistit olivat tehneet murrehaastatteluja hyvin yleisesti jo ennen kuin sosiolingvistiset vaikutteet alkoivat näkyä Suomessa 1970-luvun taitteessa. Täälläkin tutkimus uudistui, mutta murteentutkimuksen ja uuden sosiolingvistisen variaationtutkimuksen välille ei kehkeytynyt syvää juopaa (Suojanen 1982, 9; Nuolijärvi 1990, 120).

*Sosiolingvistinen haastattelu* on nimityksenä monimerkityksinen, koska sillä tarkoitetaan toisinaan mitä tahansa sosiolingvististä tutkimusta varten kasvokkaistilanteessa tallennettua puheaineistoa.[5] Joskus käsitteellä viitataan sellaiseen haastatteluproseduuriin, jollaisen William Labov (1972a; 1972b) kehitti ja joka omaksuttiin sosiolingvistiikkaan. Näiden lisäksi *sosiolingvistinen haastattelu* voi tarkoittaa jotakin edellä mainitun Labovin klassisen sosiolingvistisen haastattelun pohjalta muokattua tai sovellettua haastattelutapaa, esimerkiksi sellaista, jossa keskustelun annetaan virrata hyvin vapaamuotoisesti. (Ks. esim. Feagin 2013, 29; Becker 2013.)

Labovin (1972a; 1984) klassinen sosiolingvistinen haastattelu kytkeytyy hänen käsitykseensä tyylillisestä vaihtelusta ja siihen havaintoon, että puhujien tietoinen huomio omaan kielenkäyttöönsä vaihtelee eri

tilanteissa. Labovin New Yorkin puhekieltä tarkastelevasta tutkimuksesta lähtien sosiolingvistinen haastattelu klassisessa mielessä onkin muodostunut hyvin usein viidestä kontekstuaalisesta tyylistä. Haastatteluproseduurin aikana haastateltavalta on ollut tarkoitus saada aineistoksi valikoima informantin koko tilanteisen variaation jatkumosta, vapaamuotoisimmasta muodollisimpaan puhetapaan. (Ks. esim. Meyerhoff 2011, 34; Meyerhoff, Adachi, Nanbakhsh & Strycharz Meyerhoff 2012, 130; Becker 2013, 91–92; Meyerhoff, Schleef & MacKenzie 2015, 46–51.)

Klassinen sosiolingvistinen haastattelu muodostuu ensinnäkin varsinaisesta haastattelusta eli keskustelunomaisesta osiosta (engl. *conversational speech*), jossa haastattelija ja haastateltava(t) keskustelevat. Edellä esiteltyä terminologiaa käyttäen tässä osiossa on kyse teemahaastattelusta. Keskustelunomaisessa osiossa puhutaan etukäteen ainakin väljästi määritellyistä aihepiireistä, ja samassa yhteydessä on tavallisesti kartoitettu myös puhujan taustatietoja (esim. ikä, koulutus, ammatti), jotka ovat olleet tarpeellisia, kun on analysoitu sosiaalisten taustamuuttujien ja kielenkäytön yhteyttä.

Itse haastatteluosion lisäksi klassisessa sosiolingvistisessa haastattelussa on kolme strukturoitua elisitoitua osiota, joissa informantti joutuu kiinnittämään kieleen ja omaan puheeseensa erityistä huomiota (elisitoinneista ks. Kielten dokumentointi ja kieliopin kuvaus tk.). Ensinnäkin haastateltavalle on annettu luettavaksi teksti, jokin tarinanomainen katkelma (engl. *reading passage*). Oletuksena tässä osiossa on, että informantin kielenkäyttö on jonkin verran huolitellumpaa kuin itse haastattelussa. Tämän lisäksi informantille on annettu kaksi erillistä tehtävää, joiden oletetaan olevan epätavallisia arkiselle kielenkäytölle: sanaluetteloja (engl. *word lists*) läpikäydessä huomion oletetaan kiinnittyvän muotoon eikä niinkään merkitykseen, ja minimipareissa haastateltava tarkastelee yksittäisten sanojen välisiä pienimpiä eroja (engl. *minimal pairs*).

Kertyvästä puheaineksesta (ks. asetelma 1) varsinaisen haastattelun kielen ajatellaan muistuttavan eniten epämuodollisinta puhetapaa eli sitä, jota käytetään perheenjäsenten tai vertaisten kanssa. Tämä puheaineisto on jaettavissa Labovin mukaan (1972a; 2001) vielä tarkemmin muun muassa topiikin ja keskustelukumppanin vaihdosten perusteella tavalliseen arkipuheeseen (engl. *casual style*) ja huoliteltuun

arkipuheeseen (engl. *careful style*). (Ks. esim. Meyerhoff, Adachi, Nanbakhsh & Strycharz Meyerhoff 2012, 130; Meyerhoff, Schleef & MacKenzie 2015, 49–50.)

| | | | | |
|---|---|---|---|---|
| tavallinen arkipuhe | huoliteltu arkipuhe | luku-puhunta | sanalistat | minimiparit |

Asetelma 1. Klassisen sosiolingvistisen haastattelun viisi kontekstuaalista tyyliä epämuodollisimmasta muodollisimpaan.

Klassinen sosiolingvistisen haastattelun perusasetelma kytkeytyy tilanteiseen vaihteluun ja siihen havaintoon, että puhujien tietoinen huomio omaan kielenkäyttöönsä vaihtelee eri tilanteissa. Sen sijaan, että tutkija rakentaisi tilanteisen vaihtelun tarkastelemiseksi edellä esitellyn viisikohtaisen tutkimusasetelman strukturoituineen tehtävineen, on luonnollisesti täysin mahdollista pyrkiä tarkastelemaan kvalitatiivisesti samaista ilmiötä toisin kerätyistä vapaamuotoisista keskustelutilanteista (Meyerhoff, Schleef & MacKenzie 2015, 50). Itse asiassa Labovin käsitys edellä esitellyistä kontekstuaalisista tyyleistä on saanut osakseen runsaasti kritiikkiä: muun muassa tavallisen ja huolitellun arkipuheen erottelua toisistaan haastatteluaineistosta on pidetty ongelmallisena ja on huomautettu, että tietyissä yhteyksissä, joissa puhuja on hyvinkin tietoinen kielenkäytöstään, saatetaan puhua korostetun murteellisesti (kritiikistä ks. esim. Meyerhoff 2011, 44–49; Becker 2013, 93–94).

# 3. Haastatteluin kerätyt aineistot

Niin kuin edeltä on jo käynyt ilmi, *haastattelu* ei käsitteenä rajaa kovin tarkoin viittauskohdettaan, vaan *haastatteluksi* voidaan kutsua varsin erilaisia vuorovaikutustilanteita (ks. myös esim. 1–6). Kuka tahansa voi päätyä haastattelumaiseen keskustelutilanteeseen esimerkiksi työpaikkaa hakiessaan, lääkärin vastaanotolla tai vaikka poliisikuulustelussa.

Lehtijutut perustuvat usein journalistisiin haastatteluihin, ja lähes joka ilta voi televisiosta seurata erilaisia uutis- ja viihdehaastatteluja. Kaltaisessamme haastatteluyhteiskunnassa (ks. Silverman 1993, 248) haastattelu on paitsi suosittu myös useimmille meistä varsin tuttu ja arkinenkin puhetilanne.

Rajanveto jonkin sellaisen puhetilanteen, jota ei enää ole mielekästä pitää ensisijaisesti haastatteluna, haastattelumaisen keskustelun ja haastattelun välillä voi olla vaikeaa. Sen sijaan niin sanotun prototyyppisen haastattelun tunnusmerkit ovat selvät. Prototyyppisessä kahdenkeskisessä haastattelussa osallistujilla on ensinnäkin hieman toisistaan poikkeavat roolit, kun toinen ensisijaisesti kysyy ja kuuntelee ja toinen vastaa ja kertoo (ks. esim. 1–2; Mustanoja 2011, 82–83). Haastattelijan tehtävää hoitaa haastattelunteon ammattilainen, esimerkiksi tutkija tai toimittaja, kun taas haastateltava on valikoitunut rooliinsa, koska hänellä on hallussaan jotakin haastattelijaa kiinnostavaa: tietoa, taitoa, mielipiteitä – tai vaikka kieltä.

Esim. 1:[6]

> Haastattelija: missä päin sä oot muutem matkustellu?
>
> Informantti: (no), Ranskassa Saksassa, Itävallassa, Espanjassa Eŋlannissa, Tanskassa Ruåtsissa Norjassa. [haastattelija: joo-o] ja ooŋ käyny, Venäjäm pu(a)lella käytii. tais se oli sillon Ne$^{u}$vvostoliitto (ni).

Esim. 2:

> Haastattelija: no niin no kouluŋkäynti sitte, noil luettelomaisesti.
>
> Informantti: no, kansakoulu ja, sittet Tampereen, tyttölyseo, joka oli sillon_yhdeksäv vuoti(e), -vuotinej ja, sitte, farmasian, laitos, Helsiŋŋissä Helsiŋŋin_yliopisto$^{f}$ farmasiallaitos.

Sosiolingvistiikassa haastateltavaa on usein kutsuttu informantiksi, vanhemmassa murteentutkimuksessa kielenoppaaksi tai arvokkaasti kielimestariksi (esim. Suojanen 1982, 9; ks. myös Johdannon luku 4 tk.). Prototyyppisessä haastattelussa on läsnä aina myös ylimääräinen[7] osapuoli, mukana oleva tallennusväline. Joskus tämä tallennusväline on lehtiö ja kynä, joskus kamera ja nauhuri, ja toisinaan läsnä on kokonainen kuvausryhmä. Tallennuslaitteisiin hallintavalta on ensisijaisesti haastattelijalla, mikä sekin kuuluu prototyyppisen haastattelun roolitukseen.

Haastattelulla tuotettavaa kokemuksellista tietoa on totuttu – toisinaan jopa kritiikittömästi – pitämään jotenkin erityisen aitona ja luotettavana. Tämä on vahvistanut haastattelujen asemaa tieteellisten aineistojen keruussa. Haastattelujen suosioon on niin ikään vaikuttanut se, että haastattelu saattaa tuntua yksinkertaiselta aineistonkeruun menetelmältä. Hyvä tutkimushaastattelu ei kuitenkaan toteudu samanlaisella suoralla tekniikalla kuin toimittajan tai poliisin tekemä haastattelu. Onnistunut vuorovaikutus on paljon muuta kuin mekaanista kysymysten esittämistä. (Hyvärinen 2017, 11–12.) Haastattelun tekemisen tekniikkaan palaamme tarkemmin luvussa 4.

Haastattelusta tekee erinomaisen aineistonkeruun menetelmän kuitenkin se, että haastattelun avulla voidaan saada tietoa asioista, jotka ovat tapahtuneet aikaisemmin jossain toisaalla – tilanteessa, jossa tutkija ei ole itse voinut olla läsnä. Kielentutkimuksellisten tarkoitusperien, esimerkiksi variationistisen piirreanalyysin, näkökulmasta haastattelu myös näyttäytyy suhteellisen vapaamuotoisena puhetilanteena, jossa informantin ääntä on todennäköistä saada talteen lähes maksimaalisesti mihin tahansa muuhun vuorovaikutustilanteeseen verrattuna. On kyllä mahdollista, että esimerkiksi asiointitilanne tai ystäväjoukon peliilta tuottaa yhdeltä henkilöltä saman verran materiaalia tuntia kohden kuin haastattelu. Tutkijan vaikutusmahdollisuudet saatavan aineiston määrään ovat kuitenkin näissä tilanteissa huomattavasti vähäisemmät ja yleiset epävarmuustekijät suuremmat.

## 3.1. Tavoitteena luonnollinen, vapaamuotoinen puhe

Sosiolingvistisissä keruissa tavoitteena on tyypillisesti saada "autenttista" tai "luonnollista" puhetta, jolla tarkoitetaan väljästi tulkittuna jotain sellaista puhetapaa, jota suositaan läheisten tai ystävien kanssa, kun tilannetta ei koeta muodolliseksi (Labov 1972a, 208; Labov 1972b, 128). Tällaista puhetta nimitetään *vernakulaariksi* (engl. *vernacular*; Labov 1972b, 112; Meyerhoff, Schleef & MacKenzie 2015, 46–47). Tavoitetta on kuitenkin vaikeaa tai suorastaan mahdotonta täysin saavuttaa, sillä sosiolingvisti joutuu kohtaamaan aina havainnoijan paradoksin (engl. *observer's paradox*; Labov 1972b, 113, 209): Tutkijat ovat kiinnostuneita siitä, kuinka ihmiset puhuvat spontaanisti arkielämässään eli kuinka he puhuvat silloin, kun heidän kielenkäyttöään ei tarkkailla. Paradoksi syntyy, koska informantin puhetta halutaan nimenomaan tarkastella ja tallentaa. Tallennettaessa a) kerääjä tai tutkija on hyvin usein läsnä, b) tallennusväline – nauhuri tai kamera – on väkisinkin aina paikalla ja c) itse tallennustilanne herättää tavalla tai toisella informantin huomion (Meyerhoff, Adachi, Nanbakhsh & Strycharz 2012, 132).

*Vernakulaarikin* on käsitteenä monimerkityksinen. Yhteistä eri määritelmille on kuitenkin, että niissä vernakulaari nähdään jonkinlaisena ääripäänä tai vastakohtana kyseessä olevan kielen standardivarieteetille[8]. (Ks. esim. Milroy & Gordon 2003, 49–50; Meyerhoff 2011, 361.) Oma lukunsa on se, että tutkijat ovat kyseenalaistaneet, onko olemassa tai määriteltävissä jokin tietty luonnonmukaisin puhetapa (vernakulaari), johon tutkijat pääsisivät käsiksi, kun puhuja ei kiinnitä puheeseensa mitään huomiota. Puhujat mukauttavat kielenkäyttöään aina jotenkin, ja puhetapaan vaikuttavat muutkin seikat (esim. kommunikaation tavoitteet, sanoman sisältö ja osallistujat) kuin se, kuinka tietoisesti puhuja kiinnittää kielenkäyttöönsä huomiota. Vernakulaarin tarkastelua pidetään mielekkäänä, mutta sen yksioikoista yhdistämistä "tavalliseen puheeseen" ja normittamattomien kielimuotojen ominaispiirteisiin pidetään ongelmallisena. (Ks. esim. Milroy & Gordon 2003, 49–51; Becker 2013, 95; Schilling 2013, 60–61, 104.) Milroy ja Gordon pitävät selvänä, että variaationtutkijoiden tutkimuskohteena *vernakulaari* on käsitteenä abstraktio siinä kuin luonnollinen kielikin. Käsitteenä se on tarpeellinen

ja käyttökelpoinen, kunhan tiedostaa ja muistaa sen abstraktiuden. Tavoitteena on yleensä tarkastella luonnollista kieltä sen vapaamuotoisissa käyttöyhteyksissä, mutta koska kielenkäyttö on aina tiukasti sidoksissa käyttökontekstiin, on varottava esittämästä, että jossakin tietyssä tilanteessa informantin puhe olisi luonnollisempaa tai ”tilanteettomampaa” kuin muissa yhteyksissä. (Milroy & Gordon 2003, 50.)

## 3.2. Kritiikki ja rajoitukset

Haastatteluista on perustellusti esitetty, että ne eivät ole puhekielen tutkimuksessa ongelmaton tapa kerätä vapaamuotoista informanttien kieltä. Haastattelua pidetään arkielämälle epätavallisena genrenä, minkä vuoksi sen ei voi varauksettomasti katsoa soveltuvan täysin vapaamuotoisen puheen tallentamiseen. Toisaalta muun muassa Penelope Eckert (2000) on huomauttanut, että kielenpuhujat ovat kielellisesti sopeutuvia ja mukautuvia niin tutkimuksellisissa kuin ei-tutkimuksellisissakin kielenkäyttöyhteyksissä, minkä vuoksi he pystyvät mukauttamaan kieltään luonnollisesti (sosiolingvistiseen) haastatteluunkin sopivaksi aivan kuin kaikkiin muihin tilanteisiin, joihin he arjessaan päätyvät.

Kenties osuvin haastattelua koskeva kritiikki kohdistuu siihen, että vaikka haastattelu soveltuu erityisesti tutkimuksiin, joissa tarkastellaan taajaan puheessa esiintyviä fonologisia, morfologisia ja osin myös syntaktisia piirteitä, ovat kuitenkin useat piirteet niissä epätavallisia (esimerkiksi informantin esittämät kysymykset tai informantin käyttämä yksikön toinen persoona). Haastatteluille ominaisen rakenteen takia ne eivät sovellu esimerkiksi aineistoksi keskusteluntutkimukseen, jossa kiinnostuksen kohteena yleensä on vuorovaikutuksen rakentuminen keskustelussa (ks. esim. Milroy 1987, 24–26, 62–69; Feagin 2013, 26–27, Lappalainen 2004, 15–16). Esimerkiksi monille viime vuosisadalla tallennetuille suomalaisille murrehaastatteluille on ominaista, että äänessä niissä on enimmäkseen haastateltava. Silloisissa keruuohjeissa on usein erikseen mainittu, että koska tavoitteena on ollut tallentaa vanhakantaista murretta ja siten saada äänitettyä informantin puhetta, ei haastattelijan pidä keskeyttää puhujaa eikä lyhyiden taukojen

jälkeen pidä ryhtyä heti esittämään uusia kysymyksiä (Yli-Paavola 1970, 53). Tärkeä on huomata, että tavoitteena on ollut saada talteen toisaalta ”runsassisältöistä” ja ”vanhakantaista” mutta toisaalta ”luonnollisen vapaata puhetta” (ks. Yli-Paavola 1970, 47). Näitä kahta osatavoitetta voi pitää osin ristiriitaisina, jos pitkän monologimaisen kerronnan odottaa kattavan kaiken mahdollisen kielenkäytön luonnollisen vapaan puheen kirjosta. Ohjeistusta ja kerättyä aineistoa voi pitää kuitenkin tarkoituksenmukaisena, koska kyseistä aineistoa ei ole alun perin kerätty vuorovaikutuslingvististen aiheiden tutkimista varten. Kaiken kaikkiaan ei ole mitään sellaista keruutapaa, joka tarjoaisi soveltuvaa aineistoa kaikille tutkimuskysymyksille.

Todettava on sekin, että osa murteentutkimuksessa käytettyjen haastattelujen kritiikistä tuntuu osin kumpuavan siitä mielikuvasta, jonka *haastattelu* nimityksenä synnyttää (Kurki 2005, 65): esimerkiksi Lesley Milroy havainnollistaa perusteltua kritiikkiään esittämällä, että tiukalle haastattelulle on ominaista luokkahuonetilannetta muistuttava kysymys–vastaus-rakenne (Milroy 1987, 24–26, 62–69). Sama luonnehdinta ei ole kuitenkaan osuva, jos se kohdistetaan fennistisiin murrehaastatteluihin ja erityisesti sosiolingvistisiin haastatteluihin, sillä kuvatunlainen haastattelutapa on täkäläisissä äänitteissä epätavallinen. Tämä ilmenee kuuntelemalla äänitteitä ja tutustumalla kerääjille tehtyihin ohjeistuksiin. Sama käy ilmi myös katkelmista, jotka ovat Tampereen puhekielen haastatteluista (ks. esim. 3–6; Mustanoja 2011, 81–83, 116).

Esim. 3 (arkipäiväinen, keskustelunomainen):

> Informantti: -- ja stek kum mää oon semmone että mää lihoŋ kauhee herkästin ni em mää viitti [laittaa ruokaa], mitä mä ny ostav viiliä ja, mää söi$^{n}$ noita kanaŋkoipia yhteen_aikaam mää ostiŋ kuule että, [haastattelija: jaa.] mul_oli joka päivä kuulen niitä ja, ku$^{m}$ mää aatteli että ne ei lihota ja ne em mää stä rasvaa siittä syäny. mä ittel laitoŋ kato(k) kyä sää tiät ne koivet [joo.] sillai, pistin$^{n}$_uuniij ja, mää en_on neljääv viikkoonn_ostanu ainootaka koipee. kylä niissäkik kuulev vaan (--), ja kum mää en tiä (mutta) niistä mää kyllä varmaal lihoi. oom mää ny viis kiloo, laihtu(nu). (ku en_os syäny),

mää on_ollu$^{n}$ Nutritelle. Nutri-, mitä ne Nutri-,

Haastattelija: (--) laihdutus-, -valmisteita.

Informantti: nii joo, Nutriletti.

Haastattelija: jaa, no oŋkos, millasta se sitte oikeen o?

Informantti: no em mää tiäk ku, (em mä) oikeen tahro, viis kiloo oŋ kylä mennym mutta, mutta siin_o aikaakii kylä kulunu.

Haastattelija: miten sitä niiŋku otetaan sitä, Nutrilettia (nys sitten)?

Informantti: ne on sem(mose)sa, mää nyn näytän sulle.

Haastattelija: näytä.

Informantti: ai älä kun (ni kaŋŋ)istuu (---). [haastattelija voivottelee, informantti siirtyy kauemmas mikrofonista ja kolistelee taustalla] voi (tuolta tu$^{l}$lee) kauhee sade, [haastattelija: jaa.] (äkkiä--) tulee. nää oŋ kuulet tämmösiä pusseja. tää vaa, kylmääv veteen tää vanili-$^{j}$am makune ja sitte, stek kun_oŋ keittoo niŋ kuumaav vetteen (--). [haastattelija: jaaha, joo, joo.] tästä tu$^{l}$lee niŋkuk kaks (desilitraa).

Esim. 4 (roolien kääntyminen):

Haastattelija: no käytkös sä usein tyttären tykönä siä Helsiŋŋissä?

Informantti: juu kyllä m(ä) o, käynyn nytkin, esmes tä(n) vuodem puolela jo kaks kertaa.

Haastattelija: tulee sitten nähtyä vähän eri kaupuŋkeja tässä kuiteŋki.

Informantti: no joo kyllä, kyllä. joo. tässä, nyt$^{t}$_on_ollu yhteisiä näitä, tän, näitä haŋkkeita ni($^{n}$), niideŋki$^{n}$ takia olen_ollu.

Haastattelija: että yhteistyötä oikein.

Informantti: no joo, ei meillä ny($^{t}$), oikeestaan seŋ kummempaa o ollum mut tä- tässä nytt o(-), ihan, (sat)tu, sattu näin asioista.

[24 sekunnin hiljainen hetki]

Informantti: sulla ov Viialassa steŋ koti, m- mi-, että sää, pääset sinne, joka viikovvaihde? -vaihde ilmeisesti.

Haastattelija: no aika usein kyllä.

Informantti: tunnetko (--), Viialan, ih- ihmiset_yleensäk kaikki?

Haastattelija: no aika paljoj joo.

Informantti: tunsitko Virtasen_Erkkiä?

Esim. 5 (informantti kertoo tietyt tärkeät asiat riippumatta siitä, mitä on kysytty):

Informantti: heräs piäni, piäni tetalje, em malta ollak kertomatta niistä, minull_oli stem, vähä, minä olin, olin todella ykstoistavuatias, taikka täytin sitte heinäkuussa ykstoist(a), (ee) ja sitten_oli pikkusev vanhempia, kolmetoista ja viistoista vuatiaita ja, ja sillom me olimme siälä, Suomem paŋkin talon, kellareissa ja pitkin_aina kuv varsiŋkiŋ kun, Tampereev valtaus läheni, mutta, mutta tuata, ne sitten ne kolmetoista ja viistoista vuatiaat ne oli jo seŋ kokosia ettei ne oikeen niŋkun, arvannu, lähtee ennää, kaupuŋkille, urkkiin ja, varsiŋkin tämä viistoistavuatias veljeni, nii nii, oli kovin, sitten_aktiivinen jo, niŋkun nyt saattoi ollakkij ja, hän sitten_aina usutti että ”meen nyj ja kuuntelet tarkkaam mitä ne puhuu ja

> kuuntelen nys siälä ja ja kuuluuko Vilppular rintamalta mitääj ja mistä kuuluu ja” --

Esim. 6 (vitsailu):

> Haastattelija: käyk sää, saunavuorolla?
>
> Informantti: no mull_ei os saunavuåroo sillai kum mää käy$^{l}$ leŋkkisaunassa sitte.
>
> Haastattelija: miŋkäslainen se sitten on?
>
> Informantti: no se on semmone että kuŋ kiä(t)tää talonn_ympäri juåksee nii s$^{i}$ttev voi mennäl leŋkkisaunaa. [naurua]

Viime vuosisadan murrenauhoitteetkin ovat tavallisesti strukturoitujen haastattelujen sijasta väljästi strukturoituja tai tiukimmillaan puolistrukturoituja teemahaastatteluja. Vaikka kerääjille tehdyissä ohjeistuksissa mainitaan, että aihepiiri niissä voi olla täysin rajoittamaton, esiintyvät niissä tyypillisesti väljinä aihepiireinä haastateltavan omat elämänvaiheet ja kokemukset sekä erityisesti 1950-, 1960- ja 1970-lukujen tallenteissa kansanperinteen, kansatieteen, kansanrunouden tai paikallishistorian aiheet (Yli-Paavola 1970, 48). Useasti ohjeissa sivutaan sitä, ettei haastatteluja pidä toteuttaa kaavamaisesti, ja kerääjiä on opastettu erikseen olemaan pikemminkin tarkkaavaisia kuuntelijoita kuin kuulustelijoita ja välttämään vastausten tiukkaamista. Heitä on kehotettu sallimaan puheenaiheesta poikkeamiset ja tarpeen mukaan johdattelemaan informantit hienovaraisesti aiheeseen. Erikseen ohjeistuksessa mainitaan, että toisinaan haastattelun ollessa vapaimmillaan voi tutkija jättäytyä haastateltavan vietäväksi ja antaa informantin valita puheenaiheet. (Ks. esim. Yli-Paavola 1970, 43–45, 47.) Viimeksi mainitut ohjeet ovat sellaisia, että sellaisenaan toteutuessaan haastattelutilanne olisi ennemminkin strukturoimaton kuin strukturoitu.

Ylipäätään on kuitenkin huomattava, että pelkästään suomalaisessa murteentutkimuksessa on tehty yhteensä kymmeniä tuhansia tunteja

murreäänitteitä, ja niiden joukkoon mahtuu hyvin monenlaisia haastatteluja. *Haastattelu* on ollut Suomessa yleisnimitys usean vuosikymmenen ajan murrenauhoituksille, ja siksi pelkän nimityksen perusteella ei voi tehdä kovin pitkälle meneviä päätelmiä esimerkiksi niiden keskustelun rakenteesta, luontevuudesta tai vapaamuotoisuudesta. Kaikki suomalaisten äänitearkistojen haastatteluiksi luokitellut kielitieteelliset äänitteet (esim. SKNA, ks. tietolaatikko 1; ks. Korpusaineistot tk.) eivät ole välttämättä edes haastatteluja, vaikka ne onkin arkistoitu sillä nimellä.

## 3.3. Metodin jatkokehitys

Varsinkin 1980-luvun lopulta lähtien, jolloin puhutun kielen tutkimuksen näkökulmat ovat laventuneet entisestään (esimerkiksi keskusteluntutkimus), on haastattelumetodin rajoituksista keskusteltu kansainvälisessä ja suomalaisessa tutkimuskirjallisuudessa. Se, että haastattelun rajoituksista ja mahdollisista ongelmakohdista on keskusteltu, on johtanut siihen, että on kehitetty myös erilaisia keinoja lieventää tai ohittaa niitä (ks. esim. Milroy & Gordon 2003, 65–68). Näillä keinoilla pyritään karkeasti joko a) vaikuttamaan haastattelun sisältöön ja siihen, millaista kieltä haastattelussa esiintyy, tai b) vaikuttamaan kasvokkaishaastattelun dynamiikkaan, siis haastattelijan ja haastateltavan tai haastateltavien väliseen suhteeseen.

Eräs tällainen keino, jolla on pyritty vaikuttamaan ensin mainittuun, on ollut käytössä jo Labovin New Yorkin englantia koskevasta tutkimuksesta lähtien: siinä pyritään ohjaamaan puhe tapahtumaan, jossa haastateltava oli joutunut hengenvaaraan (ks. Labov 1972a, 93). Ajatuksena tässä oli, että tunnekuohussa ihminen ei tavallisesti kiinnitä huomiota niinkään kielenkäyttöönsä vaan asian sisältöön. Kysymällä hengenvaarasta tutkija pyrkii saamaan haastateltavan muistelemaan itse koettua kuohuttanutta tapahtumaa ja siten siirtämään mahdollisimman paljon puhujan ajatukset pois siitä, kuinka hän asiansa esittää. Keino vaikutti onnistuneelta Labovin tutkimuksessa niin kuin monissa muissakin, mutta vastaavasti useasta muusta tutkimuksesta on vastakkaisia esimerkkejä. Eri tutkijat ovat pyytäneet muistelemaan ja kertomaan

kummitusjuttuja, muistelemaan jotain merkittävää tapahtumaa lapsuudessa tai teini-ikäisiä kertomaan oman koulun eri koululaisryhmien välisistä eroista ja jännitteistä. (Milroy & Gordon 2003, 66; Feagin 2013, 29–30.) Haastattelu ja sen toteutus pitää aina sopeuttaa ja mukauttaa tarkasteltavana olevaan yhteisöön (Feagin 2013, 30). Tärkeää on yrittää miettiä ja löytää aiheita, joista haastateltavat puhuvat mielellään ja vapautuneesti.

Haastattelijan ja haastateltavan väliseen dynamiikkaan on pyritty vaikuttamaan esimerkiksi tekemällä yksilönauhoitusten sijasta pari- tai ryhmänauhoituksia. Samoin on tehty nauhoituksia, joissa tutkija ei ole ollut läsnä. Näissäkin nauhoitustavoissa on vahvuutensa ja heikkoutensa. Ensin mainituissa aineisto voi olla äänenlaadultaan heikkoa varsinkin, jos kaikilla informanteilla ei ole omia mikrofonejaan, ja sitä voi olla määrällisesti ja tyylillisesti hyvin epätasaisesti. Jälkimmäisissä on omat ongelmansa, jos keskusteltavia teemoja ja asetelmaa ei ole mietitty. Pahimmillaan tallennetut keskustelut eivät ole mitenkään keskenään vertailukelpoisia ja aineiston määrä (äänitteiden kesto, tutkittavien tapausten määrä) voi vaihdella liiaksi. Erikseen on huomattava, että vaikka tutkija ei olekaan paikalla, voivat nauhoitettavat olla korostetun tietoisia nauhoittamisesta ja puhetavastaan, mikä vaikuttaa siihen, millaista aineistoa tulee tallennetuksi. (Eri kokeiluista ja niiden hyvistä ja huonoista puolista ks. esim. Milroy & Gordon 2003, 66–67; Schilling 2013, 111–113; ks. Kielten dokumentointi ja kieliopin kuvaus; Lingvistinen etnografia toimintayhteisöissä tk.)

# 4. Haastattelun tekemisen tekniikkaa

## 4.1. Yleistä

Vaikka *haastatteluksi* voidaan kutsua varsin monenlaisia aineistonkeruutapahtumia, on haastattelujen tekemiseen mahdollista antaa joitakin yleispäteviä vinkkejä. Kokemuksen myötä kullekin sitten muodostuu oma hyväksi havaittu tyylinsä ja toimintamallinsa. Onnistuneen ja

lingvistiseen tutkimukseen soveltuvan haastatteluaineiston voi kerätä monella tavalla, mutta jos jotakin olennaista – vaikkapa videokameran lataaminen – unohtuu, tilanne saattaa olla menetetty eikä uusintahaastattelun mahdollisuutta välttämättä järjesty. Tämä koskee kaikkea lingvististä aineistonkeruuta, sillä puhetapahtumat ovat ainutkertaisia eikä niitä voi toistaa samalla tavalla kuin esimerkiksi kemiallista koetta.

Jotta lopputulos olisi mahdollisimman onnistunut, haastattelu kannattaa ajatella kolmivaiheiseksi prosessiksi, joka koostuu 1) edeltävistä toimenpiteistä, 2) varsinaisesta haastattelutilanteesta ja 3) haastattelun jälkeisistä toimenpiteistä. Vaikka itse haastattelutilanne usein ajatuksen tasolla korostuu jo pelkästään jännittävyytensä vuoksi, on syytä muistaa, että jokainen näistä kolmesta vaiheesta on oikeastaan yhtä tärkeä.

## 4.2. Ennen haastattelua

Ensimmäisessä, haastattelua edeltävässä vaiheessa huomio on kiinnitettävä informanttien valintaan ja heidän kanssaan viestimiseen, tallennusvälineisiin ja niiden käytön harjoitteluun sekä omaan valmistautumiseen. Informanttien valinta perustuu suoraan tutkimuskysymyksiin eli siihen, kenen kielenkäyttöä tai näkemyksiä tutkimuksessa selvitetään. Jos tutkimuksen kohteena on tunnetun tamperelaisen jääkiekkoilijan tremulanttiääntämys, pyritään tätä nimenomaista jääkiekkoilijaa haastattelemaan. Tällaista tutkimusta voidaan luonnehtia myös tapaustutkimukseksi. Jos taas tutkimuskohteena on yleisesti tamperelaisten tremulanttiääntämys, haastatteluun valitaan tutkimuksen laajuuden ja käytettävissä olevien resurssien rajoissa mahdollisimman edustava joukko esimerkiksi eri-ikäisiä, eri sukupuolia ja erilaisia koulutustaustoja edustavia syntyperäisiä tamperelaisia. Tässä tapauksessa tutkimuksen kohteina eli havaintoyksikköinä ovat kaikki tamperelaiset henkilöt, jotka yhdessä muodostavat tutkimuksen perusjoukon. Varmin tapa saada kattava käsitys tamperelaisten tremulanttiääntämyksestä olisi haastatella kaikki perusjoukkoon kuuluvat. Koska tällainen suurtutkimus on käytännön syistä oletettavasti mahdoton, on tutkijan muodostettava otos. Karkeasti tarkastellen informantit voidaan valita joko satunnaisotannalla

(engl. *random sample*) tai harkinnanvaraisella otannalla (engl. *judgment sample*) (otantatavoista sosiolingvistiikassa ks. esim. Schilling 2013, 31–36; Meyerhoff, Schleef & MacKenzie 2015, 37–39; ks. myös Kysely tk.).

*Edustavaksi otokseksi* voidaan kutsua sellaista tutkittavien henkilöiden joukkoa, joka kuvastaa tarkasti perusjoukon ominaisuuksia, esimerkiksi ikä-, sukupuoli- ja sosioekonomista jakaumaa. Edustava otos perustuu usein satunnaisuuteen eli haastateltavien arvontaan perusjoukosta. Aina satunnaisotanta ei kuitenkaan ole mahdollinen, välttämätön tai mielekäskään tapa hankkia henkilöitä haastateltaviksi. Esimerkiksi 1970-luvun Nykysuomalaisen puhekielen murros -hankkeen aineistonkeruussa haastatellut on kyllä valittu väestörekisterin luetteloista, mutta varsinaista satunnaisotantaa ei raportoida käytetyn. Käytännössä haastateltavat on kuitenkin valittu umpimähkäisesti ja nimet on poimittu siinä järjestyksessä kuin tutkimuksen vaatimukset täyttävä henkilö on luetteloissa esiintynyt (ks. esim. Jonninen-Niilekselä 1982, 14).

Voi myös olla, että perusjoukkoon kuuluvista havaintoyksiköistä ei edes ole olemassa tarkkaa tietoa. Esimerkkinä tällaisesta tilanteesta voidaan mainita erilaisten sosiolektien, vaikkapa homososiolektin (ks. esim. Surkka 2016, 5; Lindeman 2017, 31–34), tutkimus. Tällöin mielekäs tapa on etsiä esimerkiksi tuttavien avustuksella muutamia kohderyhmään kuuluvia henkilöitä ja pyytää heiltä haastattelun yhteydessä vinkkejä toisista tutkimukseen soveltuvista henkilöistä. Näin pikkuhiljaa muodostuu harkinnanvarainen näyte perusjoukosta. Dialektologiassa ja sosiolingvistiikassa tällaisesta aineistonhankinnan menetelmästä on käytetty osuvasti nimitystä *lumipallotekniikka*. Satunnaisotannalla muodostetun edustavan otoksen ja lumipallotekniikalla muodostetun näytteen väliin jää lisäksi lukuisia muita tapoja, joilla päteviä ja tutkimuksellisesti antoisia uusia haastatteluaineistoja voidaan hankkia. Nykypäivänä internet mahdollistaa uudella tavalla erilaisten yhteisöjen tavoittamisen – joskin muistettava on, että informanttien etsiminen internetin välityksellä rajaa sekin osan henkilöistä tutkimuksen ulottumattomiin.

Riippumatta siitä, millä tavoin haastateltavat tutkimukseen etsitään, olennaiseksi muodostuu aineistonkeruun läpinäkyvyys aina ensivaiheista lähtien. Kun tutkimuksen myöhemmälle lukijalle on selvää, millä tavoin aineisto on koostettu, myös tulosten tulkinta asettuu

mittakaavaansa. Informanttien valintaperiaatteiden ja aineistonkeruun vaiheiden merkityksen ymmärtäminen on kuitenkin tärkeintä tutkijalle itselleen: siinä missä aineistonkeruun menetelmät ovat suorassa yhteydessä tutkimuskysymyksiin,[9] tulosten tulkinta on riippuvaista tutkimusmenetelmistä – aineistonkeruun menetelmät mukaan lukien.

Toinen edustavuuteen liittyvä kysymys on, kuinka paljon informantteja tutkimuksessa on oltava, jotta aineisto olisi edustava. Usein ensimmäinen ohje tähän on se, että mitä enemmän aineistoa on, sitä parempi. Käytännöllisistä syistä on usein kuitenkin järkevää miettiä, mikä on riittävän suuri aineisto, sen sijaan, että keräisi turhan suuren aineiston, jota ei pysty kuitenkaan hallitsemaan ja analysoimaan. Kaikkiaan vastaus riippuu ennen muuta siitä, mitä tutkitaan ja mikä on tutkimuksen tavoite. (Ks. esim. Feagin 2013, 28–29; Hoffman 2014, 30–31.) Jos vertaa sitä, mitä pidetään määrällisesti edustavana otoksena esimerkiksi luonnontieteissä ja mitä kielitieteessä, on ero huomattava. Usein on esitetty, että kielitieteessä määrällisesti edustavan aineiston ei tarvitse olla niin suuri kuin luonnontieteissä (ks. tark. Milroy & Gordon 2003, 28–29 ja siinä mainitut lähteet).

Huomioon on otettava otoksen ja perusjoukon suhde ja valittava otos tarkasti niin, että se on perustellusti edustava. Jos tarkasteltaviksi halutaan ottaa esimerkiksi kielenulkoisia tekijöitä, kuten ikä, sukupuoli ja koulutus, kussakin ryhmässä on oltava riittävästi informantteja, jotta aineisto olisi edustava. On aivan eri asia, halutaanko tarkastella nykynilsiäläisten ammattikoululaisten ja lukiolaisten kielestä jotakin, ovatko tarkasteltavina vielä laajemmin pohjoissavolaiset ammattikoululaiset ja lukiolaiset vai peräti kaikki nykysuomalaiset ammattikoululaiset ja lukiolaiset. Nyrkkisääntönä esitetään usein, että viiden informantin ryhmä on riittävän edustava variationistisessa sosiolingvistiikassa (Feagin 2013, 28–29; Hoffman 2014, 30–31). Edellistä esimerkkiä soveltaen tarvitsisimme siis viisi nilsiäläistä ammattikoululaista ja viisi lukiolaista. Jos tarkastelussa on tarkoitus ottaa huomioon muitakin sosiaalisia tekijöitä, tarvitaan aineistoa enemmän. Toisin sanoen, jos haluamme tarkastella koulutuksen lisäksi myös sukupuolen vaikutusta kielenkäyttöön, tarvitsemme kaikkiaan 20 informanttia: ammattikoulusta viisi nilsiäläistyttöä ja -poikaa sekä vastaavasti lukiosta viisi tyttöä ja poikaa.

(Ks. esim. Milroy & Gordon 2003, 28–30; Feagin 2013, 28–29; Hoffman 2014, 30–31.) Sali Tagliamonte (2006, 31) esittää jopa, että kun tarkasteltavana on useita kielenulkoisia tekijöitä, riittäisi kuhunkin ryhmään kaksikin informanttia (siis esimerkiksi ammattikoulusta kaksi nilsiäläistyttöä ja -poikaa sekä kaksi lukiolaistyttöä ja -poikaa). Näin pienissä pareista muodostuvissa ryhmissä alkavat yksilölliset erot vaikuttaa jo vahvasti ryhmien ”yleisedustuksiin”. Muutenkin on tärkeää, että ryhmien muodostamisessa pyritään valikoimaan informantit niin, että he kuuluvat aidosti samaan yhteisöön ja sosiaaliseen verkostoon.

Haastateltavien valinta on vasta ensimmäinen askel kohti haastattelujen toteutusta. Seuraava askel on selvittää, suostuvatko sopiviksi katsotut haastateltavat haastatteluun, ja jos suostuvat, sopia haastattelusta heidän kanssaan. Yhteydenoton mahdollisuuksia on monia: Potentiaalisia haastateltavia voi lähestyä esimerkiksi kirjeitse, puhelimitse, tekstiviestitse, sähköpostitse tai hyödyntämällä sosiaalisen median teemoittain määrittyneitä keskusteluryhmiä tai muita ryhmiä. Joskus yhteydenotto onnistuu tuttavan tai sukulaisen avustuksella, joskus ensimmäinen kontakti on luontevimmin opettaja, työnantaja tai jonkin seuran tai yhdistyksen toimihenkilö. Olennaista on arvioida, millä tavoin yhteydenotto juuri tähän haastateltavaan tai haastateltavien joukkoon on sopivinta ja tuottaa todennäköisimmin toivotun lopputuloksen eli sen, että sopivaksi katsottu henkilö suostuu haastatteluun.

Potentiaalisten haastateltavien joukossa on aina sellaisia ihmisiä, jotka oikeanlaisen yhteydenoton seurauksena suostuvat haastateltaviksi mutta vääränlaisen yhteydenoton seurauksena kieltäytyvät. Kokemuksemme mukaan tällainen oikeanlainen yhteydenotto perustuu ennen muuta siihen, että henkilö saa heti ensikontaktin aikana asiallista ja riittävää tietoa tutkimuksesta. Olennaista on muistaa myös kertoa potentiaaliselle haastateltavalle, miten tärkeää haastateltavien löytyminen on tutkimuksen onnistumisen kannalta ja miksi juuri kyseinen henkilö soveltuisi hyvin tutkimukseen. Monille informanteille on myös tärkeää kuulla se, että haastattelutilanteessa ei ole oikeita tai vääriä vastauksia (jos tosiaan näin ei ole) vaan että kaikki vastaukset ovat yhtä arvokkaita.

Sitä, että tutkimuksessa tarkastellaan nimenomaan informantin puhekieltä tai sen tiettyä osa-aluetta, ei kuitenkaan ole syytä korostaa

yhteydenoton aikana. On havaittu, että mitä tietoisempi informantti on siitä, että hänen kieltään tarkataan, sitä tarkemmin hän kiinnittää nauhoitustilanteessa huomiota kielenkäyttöönsä ja sitä standardinomaisempaa (tai mahdollisesti liioitellun murteellista) hänen kielenkäyttönsä on. Tämä ei kuitenkaan tarkoita, että henkilölle valehdeltaisiin tai että häneltä kaikin keinoin pimitettäisiin tutkimuksen tarkoitus. Sekä tutkimuseettisesti että tutkimuksen onnistumisen kannalta mielekäs ratkaisu on kertoa tutkimuksesta esimerkiksi niin, että mainitsee yliopiston tai hankkeen, jonka piirissä tutkimusta tehdään. Tämän lisäksi tutkimusta voi luonnehtia yleisesti lingvistiseksi tai kielitieteelliseksi tai kuvata sen luonnetta yleistajuisemmin esimerkiksi niin, että tutkimuksessa kerätään aineistoa eri puolilla Suomea asuvilta ihmisiltä tai että tutkimuksessa tarkastellaan haastatteluita, joissa ihmiset kertovat aiheesta X ja Y. Jo tässä ensivaiheessa on myös hyvä jollakin tavalla luonnehtia haastattelun luonnetta (vapaamuotoinen vai strukturoitu) ja ennen muuta sen kestoa: sitä, onko kyse kymmenestä minuutista vai tunnista.

Tutkimuskysymysten ja niihin vastaamisen kannalta on olennaista miettiä paitsi haastateltavien määrää myös sitä, kuinka paljon yhdeltä haastateltavalta halutaan aineistoa. Aivan tarkkaan tätä ei voi koskaan suunnitella, sillä esimerkiksi tunnin mittainen haastattelu voi tarkoittaa tapauskohtaisesti hyvin erilaista määrää varsinaista haastattelupuhetta. Variaationtutkimuksessa tavallinen havainto on, että jotakin kielenpiirrettä voi toisella informantilla esiintyä runsaasti, kun taas toisella informantilla saman piirteen esiintymät ovat harvinaisia. Osittain ilmiössä on kyse siitä, että eri haastattelut ajautuvat erilaisille urille käsiteltävien aihepiirien mukaan. Jos esimerkiksi hämäläistaustaisen puhujan haastattelussa päädytään puhumaan *kaupunkeista* ja erityisesti *Helsinkistä* ja siellä asumiseen liittyvistä *onkelmista*, *nk*-yhtymän potentiaalisia vaihtelutapauksia saattaa esiintyä haastattelussa parikinkymmentä. Jos taas haastattelu kääntyy muihin aiheisiin, piirteen esiintymät jäävät yhden käden sormin laskettaviksi. (Ks. esim. Mustanoja 2011, 219–220, 225–226.) Jotkin variaatioilmiöt, esimerkiksi inessiivin päätteet tai diftongien avartuminen, taas ovat siinä määrin frekventtejä ja sanatasosta riippumattomia, että niiden osumisesta haastattelupuheeseen ei tarvitse ennalta murehtia.

Osittain yhden haastattelun avulla tarkasteluun saatavien piirteiden määrään vaikuttaa yksinkertaisesti osallistujien puhenopeus ja mietintätaukojen määrä ja kesto. Nämä taas palautuvat yksilöiden temperamenttiin ja omaksuttuihin vuorovaikutuskäytänteisiin. Vaikutuksensa on myös tilanteen tulkinnalla, esimerkiksi sillä, kuinka muodolliseksi tai epämuodollisen jutustelevaksi haastattelu koetaan. Kovin muodollisessa tilanteessa haastateltavan vastaukset voivat jäädä lyhyiksi ja toteaviksi, mutta myös hyvin epämuodollinen jutustelu, jossa haastattelija ottaa tilaa ja käyttää pitkiä puheenvuoroja, saattaa saada aikaan sen, että haastateltavan puhetta saadaan toivottua vähemmän. Sen sijaan oletettujen heimostereotypioiden ja saadun tutkimusmateriaalin määrän suhteesta ei ole raportoitua tutkimustietoa; ”vilkkaiden karjalaisten” ja ”hitaiden hämäläisten” joukkoon mahtuu temperamenttipiirteitä laidasta laitaan niin, että yksilövariaatio ylittää ryhmärajat monin verroin (ks. myös Yli-Paavola 1970, 56).

Haastattelijan kannalta haastattelun keston suunnittelu liittyy yleiseen ajankäytön organisointiin ja esimerkiksi sen miettimiseen, kuinka monta haastattelua yhteen päivään mahtuu. Jos haastattelu on suunniteltu esimerkiksi tunnin kestoiseksi, on haastattelijan hyvä varata sille käytännössä kaksi tuntia aikaa siirtymisiin käytettävän ajan lisäksi. Aikaa nimittäin kuluu: parkkipaikan löytämiseen, laitteiden virittelyyn, esitietolomakkeen ja tutkimuslupien täyttämiseen, suunniteltua pidemmäksi venyneeseen haastatteluun, laitteiden purkuun, kohteliaasti juotuun kahvikupilliseen – vain joitakin mainitaksemme.

Etukäteen pitää valmistautua myös tutustumalla tallennusvälineisiin ja harjoitella niiden käyttöä. Koska haastateltavien on tarkoitus kiinnittää mahdollisimman vähän huomiota tallennusvälineisiin ja haastattelutilanteeseen, pitää välttää kaikin tavoin sitä, että laitteiden toiminnan tarkistamiseen ja hapuiluun menisi aikaa. Haastattelut äänitetään nykyisin tavallisesti digitaalisin nauhurein tai videoidaan videokameran tai kannettavan tietokoneen avulla. Videokuvaamisen etuna on, että tällöin tallennetuiksi tulevat äänen lisäksi muun muassa ilmeet, eleet ja esimerkiksi, kehen katse kohdistuu tai mihin suuntaan puhuja konkreettisesti viittoo viitatessaan johonkuhun läsnäolevaan. Video helpottaa myös litterointivaihetta varsinkin niissä tilanteissa, joissa esiintyy

päällekkäispuhuntaa ja halutaan varmistua, kuka mitäkin sanoo. Siksi on tärkeää, että kamera (tai tietokone) sijoitetaan niin, että kaikki tallennettavat henkilöt näkyvät selvästi kuvassa ja että kuvanlaatu on riittävän hyvä. Kuvausteknisesti hyväksi havaittu tapa on sijoittaa haastattelijan ja haastateltavan tuolit mahdollisuuksien mukaan noin 90 asteen kulmaan toisiinsa nähden. Tällöin kamera on mahdollista sijoittaa esimerkiksi pöydälle niin, että molemmat osapuolet mahtuvat kuvaan, jolloin myös molempien ilmeet ja eleet näkyvät paremmin verrattuna esimerkiksi siihen, että haastattelun osapuolet istuutuisivat pöydän vastakkaisille puolille ja tilanne taltioitaisiin suoraan sivulta päin – tai vielä hankalammin niin, että haastateltavan kasvot kyllä näkyisivät, mutta haastattelijasta tallentuisi vain takaraivo.

Etukäteen on tutustuttava siihen, millaiselta alueelta mikrofoni tallentaa ääntä ja miten mikrofoni on syytä sijoittaa puhujiin nähden. Osa mikrofoneista tallentaa ääntä ensisijaisesti tietystä suunnasta, osa tietyistä suunnista ja osa tasaisemmin joka puolelta laitetta. Onpa kyse mistä tahansa tallennusvälineestä, on ennen haastatteluun lähtöä varmistuttava, että laitteessa on varmasti koko tilannetta varten riittävästi virtaa: jos laite ei toimi verkkovirralla, on tarkistettava akku (tai paristot). Akku- tai paristokäyttöinen laite voi tallentaa ääntä vielä silloinkin, kun virtaa on vähän jäljellä, mutta pahimmillaan se vaikuttaa äänenlaatuun niin, että toistonopeus on epätasainen.

Tallennetaanpa haastattelusta pelkkä ääni tai sekä ääni että kuva, on ehdottoman tärkeää, että äänenlaatu on muutenkin riittävän hyvä. On huomattava, että nykyään monenlaisilla puhelimilla pystyy tallentamaan ääntä, mutta vaikka korvin kuunnellen niiden äänenlaatu vaikuttaakin hyvältä, eivät ne useinkaan sovellu kielitieteellisten tallenteiden tekemiseen. Erityisen tärkeää on huomata, että jos tarkoitus on hyödyntää tallennettua aineistoa myös foneettisissa kokeellisissa mittauksissa, on digitaalisella äänitteellä minimivaatimukset. Tavallisesti tallenteet, jotka ovat 16-bittisiä ja joiden näytteenottotaajuus on 44,1 kHz, vastaavat laadultaan tutkimuksen vaatimuksia. Tämän vuoksi äänitettä ei pidä tallentaa millään sellaisella digitaalisella tallentimella, joka tallentaa äänitiedoston johonkin pakattuun formaattiin (esim. mp3): tällöin tallennuksen pakkausvaiheessa osa äänen mikrodatasta häviää, ja sitä ei

pysty enää mitenkään palauttamaan. Lähtökohtaisesti suositeltavaa on, että ääni tallennetaan esimerkiksi wav-muodossa. (ks. Kokeellinen fonetiikka tk.)

Vielä erikseen on korostettava henkisen valmistautumisen ja harjoittelun merkitystä. Jos haastattelua varten on olemassa kysymysrunko, kannattaa haastattelijan tutustua kysymyksiin huolellisesti ennen kentälle lähtöä. Ennalta tutustuminen helpottaa esimerkiksi jatko- tai täydentävien kysymysten keksimistä tilanteessa, jossa keskustelu uhkaa tyrehtyä ja aikaa olisi vielä jäljellä. Kun haastattelija tuntee haastattelurungon ja suunnitellut teemat hyvin, hän voi jo ennalta järjestellä kysymykset itselleen luontevalla tavalla. Huippu-urheilijat harrastavat mielikuvaharjoittelua, ja sama menetelmä sopisi haastattelijoillekin. Kun etukäteen miettii, miten tosipaikan koittaessa puhuu rauhallisesti, keskittyy kuulemaansa aidosti ja selvittää rennosti hankalat hetket, näin myös todennäköisemmin tapahtuu. Haastattelussa voi tapahtua mitä tahansa: puhelimet soivat, yskänkohtaus yllättää ja keskustelu ajautuu sellaisille urille, että informantti liikuttuu tai kimmastuu. Vaikeaksikin muodostuneen tilanteen saa kääntymään paremmaksi etukäteen mietityillä strategioilla. Pienille suunnistajille opetetaan ennen metsään lähtöä PJR-tekniikka: jos paniikki meinaa yllättää, **p**ysähdy **j**a **r**auhoitu. Sama pätee haastattelutilanteessa. Esimerkiksi hiljaiset hetket voivat tuntua pitkiltä silloin, kun itse haastattelijan roolissa ei keksikään sanottavaa. Jälkikäteen kuunneltuina tauot eivät kuitenkaan kuulosta lainkaan niin pahoilta.

Erittäin suositeltavaa on toteuttaa jonkinlainen harjoitushaastattelu ennen varsinaisten informanttien lähestymistä. Tämän voi toteuttaa vaikka opiskelijakaverin tai perheenjäsenen kanssa. Harjoitus ei tietenkään täysin simuloi autenttista tilannetta, mutta sen perusteella saa hyvän käsityksen haastattelurungon toimivuudesta ja siitä, miltä esimerkiksi tunnin mittainen haastattelu tuntuu. Samassa yhteydessä pääsee myös testaamaan tallennusvälineitä.

Jo ennen haastattelujen aloittamista on tärkeää huomioida tutkimuksen etiikkaan ja tietosuojaan liittyvät kysymykset. Erityisesti tietosuojaan liittyvät käytännön ohjeet hakevat vielä tämän artikkelin kirjoitushetkellä muotoaan, eikä esimerkiksi eri yliopistojen kesken ole vielä sovittu yhtenevistä käytänteistä. Muun muassa siitä, kuka opiskelijoiden tekemissä

tutkimuksissa toimii rekisterinpitäjänä, on erilaisia tulkintoja. Tämän vuoksi on hyvä keskustella käytänteistä oman työn ohjaajan kanssa tai tarvittaessa konsultoida esimerkiksi yliopiston lakimiestä. Näin tarvittavat toimet (esim. tietosuojailmoitus, haastateltavan suostumus ja riittävä informointi tutkimuksen tarkoituksesta) tulevat hoidetuiksi asianmukaisesti ja ongelmilta vältytään. Pitkään murteentutkimuksessa ja sosiolingvistiikassa katsottiin riittäväksi, että tutkija kysyy suullisen luvan informantin puheen tallentamiseen, mutta nykyään normaali käytäntö on, että informanteilta pyydetään kirjallinen suostumus. Alaikäisen haastateltavan ollessa kyseessä lupa pyydetään hänen huoltajaltaan. Useilla eri yliopistojen laitoksilla tai arkistoilla on valmiita suostumus- tai sopimuspohjia, joita voi hyödyntää. Asiallinen toiminta aineistonkeruun alkumetreiltä lähtien turvaa myös aineistojen jatkokäytön.

Haastatteluaineiston käyttöä ohjaavat tutkimuseettiset periaatteet ja pyrkimys toimia hyvän tieteellisen käytännön mukaisesti (ks. esim. TENK; Aineistonhallinnan käsikirja; ks. myös Ranta & Kuula-Luumi 2017). Aineistoa ei muun muassa saa käyttää mitenkään informanttien vahingoksi, ja sitä tulee hyödyntää niin, ettei esimerkiksi informantteja koskevia arkaluontoisia asioita tule ilmi tutkimuksista. Koska eri informantit on usein välttämätöntä pystyä erottamaan toisistaan mutta usein informanttien henkilöllisyyttä ei haluta paljastaa tai se pitää ehdottomasti salata, voidaan heille luoda pseudonyymit: esimerkiksi saman perheen tai ikäryhmän informanteille voidaan antaa samalla kirjaimella alkavat keksityt nimet tai informanteille voidaan luoda kirjaimista ja numeroista muodostuvat tunnistekoodit, jotka kertovat heidän taustastaan jotakin olennaista. Erikseen on hyvä muistaa vielä se, että tutkimusta ei voisi ylipäätään toteuttaa ilman vapaaehtoisten informanttien myötämielisyyttä, joten mahdollisuuksien mukaan on hyvä miettiä, miten tutkija pystyy vastavuoroisesti osoittamaan kiitollisuuttaan informanteille (ks. esim. Rickford 1997; Wolfram, Reaser & Vaughn 2008) ja yhteisölle vaikkapa kertomalla yhteisöä koskevan tutkimuksensa tuloksista.

## 4.3. Haastattelun aikana

Se, kuinka haastattelussa edetään ja minkä verran haastattelijan on tarkoitus ohjailla keskustelun kulkua, riippuu tutkimuksen tavoitteista ja siten myös siitä, onko kyse strukturoidusta, puolistrukturoidusta vai strukturoimattomasta haastattelusta. Kuten edellä ilmeni, on strukturoiduissa haastatteluissa yleensä tarkat aiheet ja kysymykset, minkä lisäksi kysymysten järjestys voi olla hyvin tarkasti sidottu. Teemahaastatteluissakin, joissa on usein löyhästi määritellyt aiheet, hyödynnetään erilaisia aihemoduuleja: aihepiirejä ja niihin liittyviä kysymyksiä on suunniteltu etukäteen. Jos tutkijan huomio kohdistuu keskustelun sisältöjen sijasta itse kielenkäyttöön, on näiden tarkoitus auttaa keskustelun etenemisessä. Keskustelun annetaan virrata vapaasti, mutta niitä hetkiä varten, jolloin keskustelu tuntuu tyrehtyvän, hyödynnetään samaan aihepiiriin liittyviä kysymyksiä. Etukäteen on usein mietitty myös eri vaihtoehtoisia tapoja siirtyä aihepiireistä toisiin, jotta keskustelu tuntuisi etenevän luontevasti asiasta toiseen. Aihemoduuleja hyödynsi jo Labov varhaisissa tutkimuksissaan (ks. tark. esim. Labov 1984), ja niitä on hyödynnetty myös Suomessa useissa tutkimushankkeissa ja tutkimuksissa osana haastattelurunkoa. Olennaista on, että puolistrukturoiduissa haastatteluissa aihemoduuleja ei ryhdytä käyttämään tiukkana kysymyslistana, jota seurataan jäykästi ja orjallisesti.

Jos haastattelun tarkoitus on selvittää informantin tietoja, käsityksiä tai mielikuvia määrätyistä aihepiireistä ja kysymyksistä, on aihemoduuleja suunniteltaessa mietittävä, mitkä kysymyksistä ja aiheista ovat sellaisia, joista ainakin halutaan saada tietoa. Haastateltavat – olivatpa he kuinka kiinnostuneita aiheesta ja tutkimuksesta tahansa – jaksavat vain määrällisen ajan keskittyä haastatteluun. Myös joidenkin kysymysten järjestystä voi olla tarpeen miettiä ennalta. Tyypillisesti esimerkiksi haastateltavien omaa kielenkäyttöä koskevat näkemykset on jätetty haastattelun loppuun, sillä aihepiirin esiin nostaminen saattaa vaikuttaa haastateltavan kielenkäyttöön.

Ennen muuta tärkeää on pyrkiä tekemään tallennettavasta tilanteesta mahdollisimman luonteva ja epämuodollinen. Tähän voi vaikuttaa ainakin valmistautumalla itse hyvin haastatteluun: tavoiteltavaa ei ole tehdä

haastattelussa suurta numeroa itse tilanteesta vaan pyrkiä olemaan aidosti kiinnostunut siitä, mitä toisella on kerrottavaa, ja pyrkiä reagoimaan hänen puheenvuoroihinsa. Erikseen on huomattava myös, että silläkin on merkitystä, kuinka aineistoa keräävä käyttäytyy ja kommunikoi informanttinsa kanssa ennen varsinaista haastattelua ja sen jälkeen. Mahdollinen oma jännittyneisyys tai oman kielenkäytön rekisterin muutokset vaikuttavat herkästi myös informanttiin. Haastattelun tekemiseen sisältyy usein, että informantin kanssa jutustellaan ennen tallennettavaa osiota ja sen jälkeen: jos kiirehtii töykeästi heti paikalta, kunhan on mielestään saanut riittävästi aineistoa, kiirii tutkijan huono maine pian hänen edelleen ja seuraavien informanttien löytäminen vaikeutuu. Samalla tämä voi vaikeuttaa myös toisten tutkijoiden toimintaedellytyksiä.

Koska dialektologi, sosiolingvisti tai variaationtutkija on kiinnostunut ennen muuta informanttiensa kielenkäytöstä, useissa ohjeistuksissa kehotetaan antamaan tilaa informanteille ja välttämään tilanteita, joissa kerääjä on runsaasti äänessä (esim. Feagin 2013, 31). Tutkijan liiallinen äänessä oleminen ilmenee viimeistään siinä vaiheessa, kun aineistoa litteroidaan ja havaitaan, että informantilta aineistoa on vain niukasti. Toisaalta, jotta tallennettava tilanne olisi mahdollisimman luonteva ja rento, on suositeltavaa, että kerääjä osoittaa informantille, että hän kuuntelee aktiivisesti ja on kiinnostunut siitä, mistä keskustellaan. Jotta tilanne olisi mahdollisimman keskustelunomainen, on hyvä ajatella haastattelutilannetta myös vastavuoroisuuden kannalta (ks. esim. Nordberg 1982, 108–109; Kurki 2005, 66; Feagin 2013, 31): jos oletamme, että informantit suostuvat äänitettäviksi ja kertovat usein myös henkilökohtaisista asioistaan, on odotuksenmukaista, että myös kerääjä osallistuu keskusteluun ja jakaa vastavuoroisesti kokemuksiaan ja kertoo mielipiteitään. On siis pyrittävä tasapainoilemaan kahden ääripään välillä: toisaalta vältettävä ajautumista kyselijän ja passiivisen kuuntelijan rooliin, toisaalta huolehdittava, ettei hallitse keskustelua omilla pitkillä puheenvuoroillaan ja vie tilaa informantilta.

## 4.4. Haastattelun jälkeen

Onpa haastattelu tallennettu video- tai audiolaitteistolla, on tallennettu haastattelu syytä siirtää talteen toiselle laitteelle ja tehdä siitä varmuuskopio. Varmuuskopioita ei pidä säilyttää samalla laitteella tai, jos kyse on analogisesta tallenteesta, samassa tilassa alkuperäiskappaleen kanssa. Eri yliopistoilla ja tutkimuslaitoksilla on erilaisia käytäntöjä ja palveluja aineistojen säilyttämiseen. Näihin on syytä tutustua jo etukäteen, jotta kerätty aineisto on mahdollista välittömästi siirtää asianmukaiseen ja myös tietoturvan näkökulmasta luotettavaan paikkaan. Tärkeää on varmistaa, että siirron jälkeen ylimääräisiä kopioita ei jää äänityslaitteille tai muistitikuille ajelehtimaan. Aineistojen asianmukainen säilytys on haastattelijan velvollisuus ja haastateltavan oikeus.

Mahdollisimman pian haastattelun jälkeen kannattaa kirjata muistiin haastattelusta perustiedot (aika, paikka, osallistujat, nauhoitteen kesto), joita tarvitaan ainakin, jos tallenne arkistoidaan. Hyvä on myös kirjata muistiin huomioita, vaikutelmia ja tunnelmia, joita keruuhetkellä syntyi. Tällaisia voivat olla esimerkiksi arvio haastattelun yleistunnelmasta (jännittynyt vai rento) ja raportti nauhoitustilanteen aikana sattuneista keskeytyksistä mutta yhtä lailla sekin, jos informantilla on kurkku kipeänä tai suussa juuri saadut hammasraudat artikulaatiota hankaloittamassa. Jos näitä ei kirjoita muistiin, eivät seikat palaudu mieleen enää viikkojen tai kuukausien päästä. Erityisesti jos haastattelusta on tallennettu ainoastaan ääni eikä kuvaa, myöhemmin voi olla hankalaa päästä jäljille esimerkiksi siitä, mitä kolinaa taustalla kuuluu ja kenen kolmannen henkilön ääntä mukaan on tallentunut. Mikäli muistiinpanoissa lukee, että haastateltavan puoliso kävi haastattelun loppupuolella huoneessa hakemassa kenkiään ja kertomassa lähtevänsä asioille, tilanne on selvä. Jos haastattelujen keruu muodostaa laajemman kenttäretken, voi olla tarpeen pitää varsinaista keruupäiväkirjaa.

Olennainen osa haastattelun jälkikäsittelyä on tallenteen litterointi. Litterointitapoja on useita, ja se, mitä tapaa käyttää, on ratkaistavissa sen perusteella, mitä haastatteluista tutkitaan. Puhutun kielen tutkimuksessa on karkeasti kaksi eri litterointitapaa, joista ensimmäinen on muun muassa murteentutkijoiden ja variaationtutkijoiden käyttämä tapa,

jossa pyritään tarkkaan ääntämyksen transkribointiin (esim. suomalais-ugrilainen tarkekirjoitus, SUT). Näistäkin tarkekirjoituksista on mahdollista tehdä joko karkea tai puolikarkea litterointi. (Ks. esim. Yli-Paavola 1970, 75–79; T. Itkonen 1997, 69–81; ks. myös Iivonen, Sovijärvi & Aulanko 1990, 53–60.) Muun muassa SU-transkriptiossa pyritään foneettisesti tarkkaan litterointiin, mutta sen sijaan siinä ei kiinnitetä huomiota esimerkiksi vuorovaikutuksen rakentumiseen (vierusparit, päällekkäispuhunta).

Keskusteluntutkimuksessa (CA) käytetyssä transkriptiossa kiinnitetään puolestaan huomiota muun muassa siihen, miten vuorot ja vierusparit rakentuvat, mutta sen sijaan foneettisesti tarkekirjoitus on huomattavasti suurpiirteisempää (Hakulinen & Sorjonen 1986; ks. myös Raevaara & Seppänen 1989). Monissa nykytutkimuksissa litteraatiot ovat merkintätavoiltaan jossain määrin nimenomaisen tutkimuksen tarkoituksiin räätälöityjä tai rohkeamminkin edellä esitettyä kahta perinnettä yhdistäviä versioita. Tämä käy, kunhan litteroinnin periaatteet on tutkimusaineiston esittelyn yhteydessä selkeästi kuvattu.

Litterointitavasta riippumatta on tärkeää muistaa, että olemme puhekielen emmekä litteraatioiden tutkijoita. Olipa käytössä mikä tahansa tarkekirjoitusmenetelmä, litteraatti ei koskaan voi olla täydellinen puheen kuva. Paitsi että litterointitavat eroavat toisistaan, kullakin litteroijalla on myös oma ”käsialansa”, johon vaikuttaa väistämättä myös litteroijan oma murretausta. Tyypillisesti litteroija on herkempi rekisteröimään sellaisia vivahteita, joita hänen omassa idiolektissaan ei ole. Toisen tavun vokaalinpidennykseen (esim. *lukùtoukka*; tässä tarke SUT-järjestelmän mukainen) kiinnittää herkemmin huomiota, jos ei itse ole savolaismurteiden alueelta kotoisin. Avartuneen diftongin merkinnän kriteerit taas saattavat hämäläistaustaisella tutkijalla olla tiukemmat kuin itämurteisella tutkijalla. Tämä ei tarkoita, että litterointi olisi aineiston tutkimuskuntoon saattamisen menetelmänä jotenkin kestämätön. Havainto liittyy ainoastaan siihen, että jokaisen valmiiseen litteraattiin tarttuvan tutkijan on syytä selvittää, millaisen ohjeistuksen perusteella kyseinen litteraatti on tuotettu. Jos litteroija ei ole kiinnittänyt erityistä huomiota esimerkiksi yleiskielen *d*:n vastineisiin, on tämän vaihteluilmiön tutkijan varmasti syytä kuunnella kyseiset kohdat vielä uudestaan.

Erityisen korostunut tarkistusvastuu on silloin, kun tutkimuksen tarkoituksena on vertailu ja päätelmiä tehdään esimerkiksi diakronisesta muutoksesta tai puhetilanteiden eroista. Näissä tapauksissa on aina varmistuttava, että havaitut erot ovat todellakin puhekielen eroja eivätkä eri litteroijien käsialojen eroja. (Ks. esim. Mustanoja & O'Dell 2007; Mustanoja 2011, 89.)

Litteraatti kannattaa tehdä ohjelmalla, jossa ainakin äänisignaali ja litteraatti – mahdollisuuksien mukaan myös kuva – on mahdollista kohdistaa toisiinsa. Internetistä vapaasti ladattavissa ovat muun muassa puheanalyysiohjelmat Praat ja ELAN (ks. Praat; ELAN). Näiden ohjelmien etu on erityisesti siinä, että litteraatiota tutkiessaan voi helposti kuunnella kyseisen kohdan, eivätkä varsinainen tutkimuskohde (puhe) ja litteraatti näin pääse erkanemaan toisistaan. Jos litteraatti on tehty tavallisella tekstinkäsittelyohjelmalla ja äänitiedosto on siitä täysin irrallaan, vaatii melkoisesti vaivaa etsiä aina kyseinen kohta tiedostosta kuunneltavakseen. Puheanalyysiohjelmat tuovat myös litterointivaiheeseen luotettavuutta ja sujuvuutta, sillä kuunneltavan kohdan voi ohjelmissa rajata tarkasti ja soittaa yhä uudelleen. Kokemus on osoittanut, että kuulokkeiden käyttö on litteroidessa aina ehdottoman suositeltavaa. Hyvistä laitteista huolimatta litterointi vie aikaa: yhden tunnin mittaisen äänitteen kunnolliseen transkribointiin kuluu kokeneelta tutkijalta hyvinkin työpäivän verran aikaa.

# 5. Yhteenveto: nykyiset käyttöalat ja -mahdollisuudet

Haastattelumetodia voidaan käyttää kielentutkimuksessa ainakin kolmeen tarkoitukseen: 1) ensisijaisesti puhutun kielen materiaalin tuottajana (esimerkiksi dialektologiassa, variaationtutkimuksessa ja sosiofoneettisessa tutkimuksessa), 2) ensisijaisesti informaation tuottajana (esimerkiksi nimistöntutkimuksessa ja kansanlingvistiikassa) tai 3) ensisijaisesti yhden puhetilanteen tuottajana (esimerkiksi vuorovaikutuksen

tutkimuksessa ja variaationtutkimuksessa). Ideaalitapauksessa sama haastattelu voi palvella monenlaista tutkimusta; tarkentavia litterointikierroksia toki saatetaan tarvita useita kunkin tutkimuksen tavoitteiden mukaan.

Suuri osa fennistisestä puhutun kielen aineistosta on haastattelemalla kerättyä (ks. tietolaatikot 1 ja 2). Tämä ohjaa uutta tutkimusta kahteen suuntaan: Koska haastatteluaineistoa on paljon, on mielekästä kerätä aineistoa runsaasti myös muunlaisista puhetilanteista. Vertailtavuuden ja yleisen jatkuvuuden näkökulmasta taas on ehdottoman mielekästä jatkaa myös haastattelujen tekemistä. Kuvaavimpina esimerkkeinä tästä jälkimmäisestä argumentista toimivat sosiolingvistiset reaaliaikatutkimukset, joissa koko tutkimusasetelma pohjautuu mahdollisimman samoilla periaatteilla koostettujen eriaikaisten aineistojen vertailuun. Esimerkiksi useissa suomalaisissa haastatteluin aloitetuissa pitkittäisseuruissa keruuperiaatteiden muuttaminen kesken kaiken ei tulisi kysymykseen, sillä se romuttaisi tutkimuksen perustan – aineiston vertailtavuuden (ks. esim. Kurki 2005; Palander 2005; Paunonen 2005; Nuolijärvi & Sorjonen 2005; Mustanoja 2011).

Haastattelut tarjoavat perinteisten käyttöalojensa lisäksi mahdollisuuksia myös uusiin tutkimuksellisiin avauksiin (ks. myös Uusitupa 2017, 78). Haastattelujen tyypillisesti hyvä äänenlaatu tarjoaa edellytykset esimerkiksi foneettisiin mittauksiin ja sen myötä vahvemmin sosiofoneettiseen tutkimusotteeseen. Henkilöhaastatteluissa toistuvat samankaltaiset aihepiirit taas mahdollistavat monenlaisen vertailun ja tarjoavat materiaalia esimerkiksi diskurssintutkimuksen tarpeisiin. Muun muassa kansanlingvistisessä tutkimuksessa haastattelut ovat saaneet vahvan roolin informaation tuottajina ja tarjonneet kosketuspintaa puhujien käsitysten ja heidän todellisen kielenkäyttönsä suhteeseen. Haastattelujen monipuolisen ja uudenlaisen käytön esteeksi voi oikeastaan muodostua vain liian kapea tai stereotypioihin rajoittunut kuva haastattelemalla kerätyn aineiston luonteesta.

Se, että haastatteluita on käytetty jo varhaisessa dialektologisessa tutkimuksessa, saattaa virheellisesti johtaa ajatukseen haastattelusta ei pelkästään vanhana vaan myös jollakin tapaa vanhentuneena aineistonkeruun metodina. Menetelmän pitkäaikainen käyttö on kuitenkin

ainoastaan sen etu: haastatteluiden kohdalla olemme jo varsin hyvin perillä siitä, millaista materiaalia niillä saadaan kerättyä, millaisiin lingvistisiin tutkimuksiin ne erityisesti sopivat ja mitkä taas ovat kielentutkimuksen näkökulmasta menetelmän rajoitukset. Hyvällä tutkimusmenetelmällä ei ole parasta ennen -päiväystä, eikä se kulu käytössä. Silti hyvääkin menetelmää kannattaa säännöllisesti pysähtyä tarkastelemaan, ja sitä kannattaa aina myös pyrkiä kehittämään.

**TIETOLAATIKKO 1.**

**Suomen kielen nauhoitearkisto (SKNA)**

- perustettu vuonna 1959, nykyisin osa Kotimaisten kielten keskusta (Kotus)
- noin 24000 tuntia kielitieteellisiä äänitteitä suomen murteista, suomen sukukielistä ja muista kielistä sekä kulttuurihistoriallisia äänitteitä esitelmistä, akateemisista tilaisuuksista ja henkilöhaastatteluista
- arkiston aineistoja käyttävät kielentutkijoiden ja kieliaineiden opiskelijoiden lisäksi myös kulttuurintutkijat, murteiden ja paikallishistorioiden harrastajat sekä äänitettyjen kielenoppaiden jälkeläiset
- kaksi kolmasosaa aineistosta digitoitu (tilanne marraskuussa 2017)
- osa aineistosta käytettävissä myös Kielipankissa, jossa äänitteet ja niistä tehdyt litteroinnit on kohdistettu, eli niitä pystyy seuraamaan samanaikaisesti.

https://www.kotus.fi/aineistot/puhutun_kielen_aineistot.

**TIETOLAATIKKO 2.**

**Lauseopin arkisto (LA)**

- perustettu valtion humanistisen toimikunnan kanssa tehdyllä sopimuksella vuonna 1967 samaan aikaan Muoto-opin arkiston kanssa
- Lauseopin arkisto sijoitettu Turun yliopistoon, Muoto-opin arkisto Helsingin yliopistoon

- suomen kielen ensimmäinen sähköinen korpus, joka koostuu kaikkiaan 132:sta morfologisesti ja syntaktisesti koodatusta aineistoyksiköstä ja aineiston käytön mahdollistavasta ohjelmasta
- 132 yksikköä murteita
- kestoltaan noin tunnin mittaisista äänitteistä tehdyt litteroinnit, jotka on annotoitu morfologisesti ja syntaktisesti
- käytettävissä Kielipankissa
- 2000-luvulla Lauseopin arkiston alaisuuteen on liitetty useita muita sähköisiä annotoituja ja annotoimattomia suomen kielen korpuksia.

**Aiheesta lisää:**

Hyvärinen, Matti, Nikander, Pirjo & Ruusuvuori, Johanna. 2017. *Tutkimushaastattelun käsikirja*. Tampere: Vastapaino.

Meyerhoff, Miriam, Adachi, Chie, Nanbakhsh, Golnaz & Strycharz, Anna. 2012. Sociolinguistic fieldwork. Julkaisussa: Thieberger, Nicholas (toim.) *The Oxford Handbook of Linguistic Fieldwork*. Oxford: Oxford University Press, 122–146.

Meyerhoff, Miriam, Schleef, Erik & MacKenzie, Laurel. 2015. *Doing Sociolinguistics. A Practical Guide to Data Collection and Data Analysis*. New York (NY): Routledge.

Schilling, Natalie. 2013. *Sociolinguistic Fieldwork. Key Topics in Sociolinguistics*. Cambridge: Cambridge University Press.

## VIITTEET

1 Tutkimuskirjallisuudessa eri haastattelulajien nimitykset vaihtelevat. Strukturoimatonta haastattelua kutsutaan muun muassa *avoimeksi haastatteluksi* ja *standardoimattomaksi haastatteluksi*. Strukturoitua haastattelua on puolestaan nimitetty esimerkiksi *standardoiduksi haastatteluksi* ja puolistrukturoitua *puolistandardoiduksi haastatteluksi* (Tiittula & Ruusuvuori 2009, 11–12). Tutkijoiden keskuudessa ei ole yhtenäistä tapaa määritellä niitä haastattelun lajeja, jotka sijoittuvat strukturoitujen ja strukturoimattomien haas-

tattelujen väliin. Esimerkiksi joidenkin tutkijoiden mukaan näiden ääripäiden väliin ovat erotettavissa osittain järjestelty ja osittain avoin haastattelu. (Hirsjärvi & Hurme 2000, 49; Saaranen-Kauppinen & Puusniekka 2006.)

2 Sosiolingvistiikassa haastattelut ovat olleet keskeinen aineistonhankintamenetelmä 1970-luvulta lähtien. Tässä artikkelissa esitettäviä metodologisia perusperiaatteita on nähdäksemme mahdollista soveltaa monella muullakin lingvistisen tutkimuksen alalla.

3 Vaihtoehtoisesti tässä yhteydessä voisi käyttää myös nimitystä *murremaantiede* kuten Chambers ja Trudgill (1998, 14) esitellessään perinteisen murteentutkimuksen keruumenetelmiä.

4 Esimerkiksi suomalaistutkijat pyrkivät kirjaamaan sanelun mukaan satu- ja tarinanäytteitä ja varsinaisten kirjattavien kielenainesten lähikontekstia.

5 Esim. Kara Becker (2013) erottelee sosiolingvististä haastattelua esitellessään tiukan perinteisen merkityksen (engl. *the sociolinguistic interview*) ja väljän merkityksen (engl. *sociolinguistic interview*).

6 Haastatteluesimerkkien litteroinnista ks. Mustanoja 2011, 86–87. Informanttien nimet ja muut henkilötietoihin viittaavat yksityiskohdat on tässä muutettu samoin kuin alkuperäislähteessä.

7 Ylimääräinen esimerkiksi suhteessa ns. observoijan paradoksiin (Labov 1972a, 113, 209).

8 *Standardilla* tarkoitetaan kyseessä olevan kielen normitettua kielimuotoa. Esimerkiksi suomen kielellä puhutaan ja kirjoitetaan eri tavoilla, mutta sen kirjakielellä on standardin asema.

9 Silloinkin, kun hyödynnetään jo olemassa olevaa aineistoa, tutkimuskysymysten, -metodien ja -tulosten keskinäinen suhde on samanlainen.

## KIRJALLISUUS


Aineistonhallinnan käsikirja. Tampere: Yhteiskuntatieteellinen tietoarkisto [ylläpitäjä ja tuottaja]. [verkkoaineisto]. [viitattu 14.12.2017]. Saatavissa: http://www.fsd.uta.fi/aineistonhallinta/.

Bamberg, Jaakko. 2017. Kävelyhaastattelu. Julkaisussa: Hyvärinen, Matti, Nikander, Pirjo & Ruusuvuori, Johanna (toim.) *Tutkimushaastattelun käsikirja*. Tampere: Vastapaino, 256–269.

Becker, Kara. 2013. The sociolinguistic interview. Julkaisussa: Mallinson, Christine, Childs, Becky & Herk, Gerard Van (toim.) *Data Collection in Sociolinguistics*. New York (NY) & London: Routledge, 91–100.

Chambers, J. K. & Trudgill, Peter. 1998. *Dialectology*. (2. p.). Cambridge: Cambridge University Press.

Eckert, Penelope. 2000. *Linguistic Variation as Social Practice*. Oxford: Blackwell.

ELAN. ELAN – The Language Archive. [verkkoaineisto]. [viitattu 15.12.2017]. Saatavissa: https://tla.mpi.nl/tools/tla-tools/elan/.

Feagin, Crawford. 2013. Entering the community. Julkaisussa: Chambers, J. K. & Schilling, Natalie (toim.) *The Handbook of Language Variation and Change*. (2. p.). Oxford: Wiley, 19–37.

Hakulinen, Auli & Sorjonen, Marja-Leena. 1986. Keskustelun kuvaamisesta. Julkaisussa: Leino, Pentti & Kalliokoski, Jyrki (toim.) *Kieli 1*. Helsinki: Helsingin yliopiston suomen kielen laitos, 5–38.

Hirsjärvi, Sirkka & Hurme, Helena. 2000. *Tutkimushaastattelu. Teemahaastattelun teoria ja käytäntö*. Helsinki: Yliopistopaino.

Hoffman, Michol. 2014. Sociolinguistic interviews. Julkaisussa: Holmes, Janet & Hazen, Kirk (toim.) *Research Methods in Sociolinguistics. A Practical Guide*. Oxford: Wiley, 25–41.

Hyvärinen, Matti. 2017. Haastattelun maailma. Julkaisussa: Hyvärinen, Matti, Nikander, Pirjo & Ruusuvuori, Johanna (toim.) *Tutkimushaastattelun käsikirja*. Tampere: Vastapaino, 1–45.

Hyvärinen, Matti, Nikander, Pirjo & Ruusuvuori, Johanna. 2017. *Tutkimushaastattelun käsikirja*. Tampere: Vastapaino.

Iivonen, Antti, Sovijärvi, Antti & Aulanko, Reijo. 1990. *Foneettisen kirjoituksen kehitys ja nykytila. Kansainvälinen foneettinen aakkosto (IPA). Suomalais-ugrilainen tarkekirjoitus (SUT)*. Helsingin yliopiston fonetiikan laitoksen monisteita n:o 16. Helsinki: Helsingin yliopisto.

Ikonen, Hanna-Mari. 2017. Puhelinhaastattelu. Julkaisussa: Hyvärinen, Matti, Nikander, Pirjo & Ruusuvuori, Johanna (toim.) *Tutkimushaastattelun käsikirja*. Tampere: Vastapaino, 270–286.

Itkonen, Terho. 1997. *Nimestäjän opas.* (3., uud. p.). Helsinki: Suomalais-Ugrilainen Seura.

Jonninen-Niilekselä, Kaija. 1982. Tampereen puhekieli tutkimuskohteena. Julkaisussa: Jonninen-Niilekselä, Kaija (toim.) *Tampereen puhekieli tutkimuskohteena*. Folia Fennistica & Linguistica. Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitoksen julkaisuja 6. Tampere: Tampereen yliopisto, 7–35.

Juusela, Kaisu. 1994. Variation as a theme in Finnish dialectology and sociolinguistics in recent years. Julkaisussa: *Fenno-ugrica Suecana 12*. Tidskrift för finsk-ugrisk forskning i Sverige. Finska institutionen. Stockholm: Stockholms universitet, 1–54.

Kurki, Tommi. 2005. *Yksilön ja ryhmän kielen reaaliaikainen muuttuminen. Kielenmuutosten seuraamisesta ja niiden tarkastelussa käytettävistä menetelmistä*. SKST 1036. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kurki, Tommi & Mustanoja, Liisa. 2019. Käänteitä, aaltoja ja risteymiä fennistisessä variaationtutkimuksessa. Julkaisussa: Hamunen, Markus, Nieminen, Tommi, Kelomäki, Tapani & Dufva, Hannele (toim.) *Käänteitä ja käsitteitä*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Labov, William. 1972a. *Sociolinguistic Patterns.* Philadelphia: University of Pennsylvania Press.

Labov, William. 1972b. Some principles of linguistic methodology. *Language in Society* 1, 97–120.

Labov, William. 1984. Field methods of the project on linguistic change and variation. Julkaisussa: Baugh, J. & Sherzer, J. (toim.) *Language in Use: Readings in Sociolinguistics.* Englewood Cliffs (NJ): Prentice Hall, 28–66.

Labov, William. 2001. The anatomy of style-shifting. Julkaisussa: Eckert, P. & Rickford, J. (toim.) *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Lappalainen, Hanna. 2004. *Variaatio ja sen funktiot. Erään sosiaalisen verkoston jäsenten kielellisen variaation ja vuorovaikutuksen tarkastelua*. SKST 964. Helsinki: Suomalaisen Kirjallisuuden Seura.

Lindeman, Meri. 2017. Homoässä, hintonaatio ja heiluvat kädet – kansanlingvistisiä käsityksiä suomenkielisten homomiesten puheesta. Pro gradu -tutkielma. Viestintätieteiden tiedekunta. Suomen kielen tutkinto-ohjelma. Tampereen yliopisto.

Meyerhoff, Miriam. 2011. *Introducing Sociolinguistics*. Oxon: Routledge.

Meyerhoff, Miriam, Adachi, Chie, Nanbakhsh, Golnaz & Strycharz, Anna. 2012. Sociolinguistic fieldwork. Julkaisussa: Thieberger, Nicholas (toim.) *The Oxford Handbook of Linguistic Fieldwork*. Oxford: Oxford University Press, 122–146.

Meyerhoff, Miriam, Schleef, Erik & MacKenzie, Laurel. 2015. *Doing Sociolinguistics. A Practical Guide to Data Collection and Data Analysis*. New York (NY): Routledge.

Mielikäinen, Aila. 1982. Nykypuhesuomen alueellista taustaa. *Virittäjä* 86, 277–294.

Milroy, Lesley. 1987. *Observing and Analyzing Natural Language. A Critical Account of Sociolinguistic Method*. Language in Society 12. Oxford: Blackwell.

Milroy, Lesley & Gordon, Matthew. 2003. *Sociolinguistics. Method and Interpretation*. Language in Society 34. Oxford & Berlin: Blackwell.

Mustanoja, Liisa. 2011. *Idiolekti ja sen muuttuminen. Reaaliaikatutkimus Tampereen puhekielestä*. Acta Universitatis Tamperensis 1605. Tampere: Tampereen yliopisto.

Mustanoja, Liisa & O'Dell, Michael. 2007. Suomen *d* ja *r* sosiofoneettisessa kentässä. *Virittäjä* 111, 56–67.

Nordberg, Bengt. 1982. Sociolingvistisk datainsamling. Julkaisussa: Thelander, Mats (toim.) *Talspråkforskning i Norden, Mål – material – metoder*. Lund: Studentlitteratur, 92–119.

Nuolijärvi, Pirkko. 1990. Keskusteluntutkimus. Julkaisussa: Mäkelä, Klaus (toim.) *Kvalitatiivisen aineiston analyysi ja tulkinta*. Helsinki: Gaudeamus, 114–141.

Nuolijärvi, Pirkko & Sorjonen, Marja-Leena. 2005. *Miten kuvata muutosta? Puhutun kielen tutkimuksen lähtökohtia murteenseuruuhankkeen pohjalta*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 133. Helsinki: Kotimaisten kielten tutkimuskeskus.

Palander, Marjatta. 2005. *Lapsuudesta keski-ikään. Seuruututkimus itäsavolaisen yksilömurteen kehityksestä*. Suomi 191. Helsinki: Suomalaisen Kirjallisuuden Seura.

Paunonen, Heikki. 2005. Helsinkiläisiä puhujaprofiileja. *Virittäjä* 109, 162–200.

Praat. Praat: Doing Phonetics by Computer. [verkkoaineisto]. [viitattu 15.12.2017]. Saatavissa: http://www.fon.hum.uva.nl/praat/.

Raevaara, Liisa & Seppänen, Eeva-Leena. 1989. Litterointi puheen kuvana ja tekstinä. Julkaisussa: Hakulinen, Auli (toim.) *Suomalaisen keskustelun keinoja I*. Kieli 4. Helsinki: Helsingin yliopiston suomen kielen laitos, 29–36.

Ranta, Juha & Kuula-Luumi, Arja. 2017. Haastattelun keruun ja käsittelyn ABC. Julkaisussa: Hyvärinen, Matti, Nikander, Pirjo & Ruusuvuori, Johanna (toim.) *Tutkimushaastattelun käsikirja*. Tampere: Vastapaino, 413–426.

Rickford, John Russell. 1997. Unequal partnership. Sociolinguistics and the African American speech community. *Language in Society* 26, 161–198.

Ruusuvuori, Johanna & Tiittula, Liisa (toim.). 2009. *Haastattelu. Tutkimus, tilanteet ja vuorovaikutus*. (2. p.). Tampere: Vastapaino.

Ruusuvuori, Johanna, Nikander, Pirjo & Hyvärinen, Matti (toim.). 2010. *Haastattelun analyysi*. Tampere: Vastapaino.
Saaranen-Kauppinen, Anita & Puusniekka, Anna. 2006. *KvaliMOTV – Menetelmäopetuksen tietovaranto*. Tampere: Yhteiskuntatieteellinen tietoarkisto [ylläpitäjä ja tuottaja]. [verkkoaineisto]. [viitattu 8.11.2017]. Saatavissa: http://www.fsd.uta.fi/menetelmaopetus/.
Schilling, Natalie. 2013. *Sociolinguistic Fieldwork*. Key Topics in Sociolinguistics. Cambridge: Cambridge University Press.
Silverman, David. 1993. *Interpreting Qualitative Data. Methods for Analyzing Talk, Text and Interaction*. London: Sage.
Suojanen, Matti K. 1982. Sosiolingvistiikka – synkretistinen tieteenala. Julkaisussa: Suojanen, Matti K. & Suojanen, Päivikki (toim.) *Sosiolingvistiikan näkymiä*. Helsinki: Gaudeamus, 7–37.
Strandberg, Jan. 2004. Ei sanat salahan joua. Fennistiikan murteenkeruun historiaa 1868–1925. Pro gradu -tutkielma. Helsingin yliopisto, suomen kielen laitos.
Surkka, Sanni. 2016. Tyylikästä lässytystä. Homoseksuaali mies kielellisenä stereotyyppinä ja identiteettinä. Pro gradu -tutkielma. Helsingin yliopisto, suomen kielen, suomalais-ugrilaisten ja pohjoismaisten kielten ja kirjallisuuksien laitos.
Tagliamonte, Sali. 2006. *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.
TENK = Tutkimuseettinen neuvottelukunta (TENK). Saatavissa: https://www.tenk.fi.
Tiittula, Liisa & Ruusuvuori, Johanna. 2009. Johdanto. Julkaisussa: Ruusuvuori, Johanna & Tiittula, Liisa (toim.) *Haastattelu. Tutkimus, tilanteet ja vuorovaikutus*. (2. p.). Tampere: Vastapaino, 9–21.
Törrönen, Jukka. 2017. Virikehaastattelu. Julkaisussa: Hyvärinen, Matti, Nikander, Pirjo & Ruusuvuori, Johanna (toim.) *Tutkimushaastattelun käsikirja*. Tampere: Vastapaino, 233–255.
Uusitupa, Milla. 2017. *Rajakarjalaismurteiden avoimet persoonaviittaukset.* Publications of the University of Eastern Finland. Dissertations in Education, Humanities, and Theology, No 117. Joensuu: University of Eastern Finland.
Wolfram, Walt, Reaser, Jeffrey & Vaughn, Charlotte. 2008. Operationalizing linguistic gratuity. From principle to practice. *Language and Linguistics Compass* 2/6, 1109–1134.
Yli-Paavola, Jaakko. 1970. *Vuosikymmen kielennauhoitusta. Suomen kielen nauhoitearkiston toimintaa v.* 1959–1968. Helsinki: Suomalaisen Kirjallisuuden Seura.

# Luku 4 Kokeellinen fonetiikka

*Michael O'Dell*
https://orcid.org/0000-0001-6496-2805

*Tommi Nieminen*

**Mikä?**

Fonetiikka tutkii puhetta ja puheen ominaispiirteitä. Sitä voi pitää osana kielitiedettä tai aivan omana tieteenalanaan. Koska kaikkia ihmisen tuottamia ääniä ei pidetä puheena vaan ainoastaan kielellistä merkitystä kantavia ääniä, tutkimusalueen rajat riippuvat kielelle ja merkityksellisyydelle määritetyistä rajoista. Itse tutkimuskohdettaan, puhetta, fonetiikka tutkii kaikin mahdollisin keinoin. Perinteisessä kolmijaossa keinot jakautuvat akustisiin (itse äänisignaalia koskeviin), artikulatorisiin (äänen tuottamista koskeviin) ja auditiivisiin tai perseptuaalisiin (äänen havaitsemista koskeviin) menetelmiin.

**Katso myös:**

Johdannon luvut 3.1. Kieltä ei ole -teesi ja 3.2. Ontologia ja kielentutkimus

Luku 24 Psyko- ja neurolingvistiikka

# 1. Johdanto

Fonetiikkaan voi suhtautua niin kielitieteen yhtenä osa-alueena kuin omana itsenäisenä tieteenalanaan. Siinä tutkitaan puheen ja sitä tuottavien ja vastaanottavien järjestelmien yksityiskohtia. Puhe voi olla äänellistä tai viitottua. Tutkittavia yksityiskohtia ovat mm., mistä osista puhe koostuu, miten osat kytkeytyvät toisiinsa, mitä puhujassa ja hänen puhe-elimistössään tapahtuu puhetta tuotettaessa sekä miten kuulija vastaavasti havaitsee puheen ja käsittelee havaintoaan. Näitä seikkoja on kaiken kaikkiaan niin paljon, ne ovat niin monenlaisia ja vaativat niin erilaisia tutkimustapoja, että fonetiikan voi luontevasti katsoa myös kielitieteestä erilliseksi tieteenalaksi.

Kielitiedettä ja fonetiikkaa yhdistävänä tekijänä on siis puhe, mutta fonetiikan ja kielitieteen näkökulma siihen voi olla hyvin erilainen. Jo kielitieteessäkin *puheen* käsitteen voi nähdä monin eri tavoin, mutta useimpia tapoja yhdistää kuitenkin ajatus puheesta yhtenä kielen muotona tai kielen konkreettisena toteutumana. Fonetiikan tutkijat sen sijaan eivät välttämättä sijoita kieltä asetelmaan ollenkaan (ainakaan millään yksinkertaisella tavalla): puhetta voidaan tarkastella omana itsenäisenä tutkimuskohteenaan ja sen mahdollistavia ja toteuttavia järjestelmiä osana nimenomaan puheen, ei abstraktimmin kielen, teoreettista ulottuvuutta. Yksi seuraus tästä on, että fonetiikka on kauttaaltaan sangen empiirinen tiede, jopa teoreettisimmillaan, ja muistuttaa siten usein pitkälti luonnontiedettä. Yksi tapa ajatella fonetiikkaa onkin nähdä sen kuuluvan niin ihmistieteisiin kuin luonnontieteisiin ja yhdistävän kummankin ajattelu- ja toimintatapoja.

Suhteessa moniin kielitieteen menetelmiin fonetiikan menetelmät ovat usein hyvinkin invasiivisia. Tavat, joilla foneettista aineistoa kerätään, vaativat monesti tutkijaa tulemaan hyvin lähelle niitä ihmisiä, joiden puhetta tai puheentuottamis- ja vastaanottamisjärjestelmää tutkitaan – ”tunkeutumaan” heidän normaaliin olotilaansa ja puhetoimintaansa. Tutkija ei juurikaan pääse vetäytymään pois tilanteesta, jossa hän ja tutkittava ovat hyvin konkreettisesti läsnä. Tämä pätee jopa akustiseen fonetiikkaan (ks. luku 3.2.), koska foneettinen tutkimus asettaa

kerättävälle ääniaineistolle usein paljon tarkemmat laatuvaatimukset kuin ”tyypillinen” kielitieteellinen tutkimus.

Fonetiikan menetelmien invasiivisuus osaltaan selittää, miksi tieteenalaan on ajoin kohdistettu syytöksiä laboratoriotutkimuksesta eli tutkimuksesta, jonka aineistona ei ole todellinen kielenkäyttö. Foneetikon voi olla kielitieteilijääkin vaikeampi torjua observoijan paradoksi eli tutkimuksen vaikutus tutkittavaan. Fonetiikka ei tieteenalana niinkään väen väkisin ”pyri” laboratorioon, vaan sen on tutkimusmenetelmiensä ja tutkimuksen toistettavuuden takia sinne usein pakko mennä. Samasta syystä fonetiikka on usein jäänyt kuvaamaan monologista puhetta. Monologisuus ei ole osa fonetiikan kuvaa puheen luonteesta, mutta sitä on tutkittu dialogisia tilanteita enemmän tutkimusmenetelmällisistä syistä. Tutkimusalaan kuuluu sinänsä kaikki puhe, myös vuorovaikutteiset puhetilanteet.

Fonetiikalla on runsaasti eri menetelmiä, mikä seuraa luonnollisesti sekä tutkimuskohteen moninaisuudesta että vaikeudesta tutkia sen muotoja. Tässä keskitytään esimerkinomaisesti menetelmistä vain joihinkin.

# 2. Kokeellisen fonetiikan historiaa ja taustaa

## 2.1. Historiaa

Vaikka fonetiikka ja kielitiede on tieteen metodologiassa erotettu eri aloiksi vasta myöhään, 1800–1900-lukujen taitteessa, fonetiikkaa on tutkittu yhtä pitkään kuin kieltä muutenkin. Tämä on tietysti luonnollista. Jos kieltä ylipäänsä aikoo tutkia, on pakko ottaa huomioon, että valtaosa siitä syntyy (edelleenkin) puheena ja että puhe on ensisijainen modaliteetti niin yksilön- kuin lajinkehityksessäkin. Sanojen muoto rakentuu ja merkityksiä välitetään niin usein äänteellisesti, ettei äänteellistä ulottuvuutta voi kokonaan ohittaa.

Tietyssä mielessä voi itse asiassa väittää, että varhaisen (soveltavan) fonetiikan tutkimuksen merkittävin kehitysaskel on ollut puhutun

kielen muotoon perustuvien kirjoitusjärjestelmien – kuten käyttämämme latinalaisen kirjoituksen – kehitys. Hyvin pitkään kirjoitetun kielen antama malli ohjasi tutkimusta jopa siinä määrin, että yhä edelleen monet naiivit kielenpuhujat pitävät kirjainta ja äännettä samamerkityksisinä; esimerkiksi lapsesta sanotaan usein, että ”hän ei vielä osaa *kirjainta s*”, vaikka lapsi ei osaisi kirjoittaa lainkaan. Sama kirjainten malliin uskominen näkyy myös suuressa osassa nykyaikaista kielitiedettä (ks. esim. Linell 2005).

Tieteenalan historiassa tekniikan kehityksellä on ollut suuri merkitys (ks. esim. Ashby 2017). Monesti uusia fonetiikan menetelmiä on syntynyt, kun aivan muuta päämäärää varten luotua tekniikkaa on keksitty soveltaa tutkimustarkoituksissa puheeseen. Hyvin suuri rooli on ollut esimerkiksi puhelintekniikan kehitystyöllä niin vanhojen lankapuhelinten kuin nykyisten kännyköidenkin aikana. Koska puhelimitse siirrettäessä ääni on yksinkertaisesti bittivirta, operaattorien on tavattoman tärkeä tietää, mikä datassa on oleellista ja mikä on mahdollista ohittaa viestinnän laadun merkittävästi kärsimättä. Puheäänen akustisten ja sen havaitsemisen auditiivisten ominaisuuksien tunteminen antaa esimerkiksi mahdollisuuden pakata ääntä mahdollisimman tehokkaasti niin, että kaikki oleellinen signaalista kuitenkin välittyy, eivät pelkästään sanat vaan myös vaikkapa tieto puhujan identiteetistä.

Jo varhaisimmissa kielitieteellisissä kuvauksissa kielen äänellinen puoli olikin läsnä. Länsimaiden antiikissa tutkimusta rajoitti ehkä eniten kielitieteen varsin filosofinen tutkimusote. Empiiristen menetelmien tai empiriaa koskevien havaintojen saatettiin kokea jopa loukkaavan kielen arvoa ihmisen ominaisimpana erityispiirteenä sekä ajattelun ja ajatusten välittämisen tärkeimpänä välineenä.

Pitkään ei erotettukaan toisistaan kielen fonologiaa (merkitystä erottavien äänteiden tasoa) tai puheen fonetiikkaa (kaikkia puheeseen liittyviä ilmiöitä siitä riippumatta, erottavatko ne määräkielessä merkitystä tai eivät). Osaltaan tähän vaikutti kirjoituksen käyttäminen puheen mallina. Aivan kuten kirjoituksen myös puheen katsottiin koostuvan ”kirjaimista”, toisin sanoen äänteistä, jotka tuotetaan yksi kerrallaan ja jotka seuraavat toisiaan enimmäkseen erillisinä segmentteinä, lomittumattomina sekventiaalisina elementteinä. Nykyään kuitenkin tiedetään,

että puhetta tuotettaessa puhe-elimistön tapahtumat paitsi lomittuvat myös mukautuvat toisiinsa jopa niin paljon, että joskus ne sulautuvat ajallisesti kokonaan yhteen. 1900-luvun loppupuolelta alkaen on myös yhä selvemmin tajuttu, että kirjoitus jättää valtaosin merkitsemättä yhden puheen keskeisistä toimintarakenteista, prosodian, johon kuuluvat yleisen näkemyksen mukaan ainakin sävelkulku, kestot ja rytmi tai painotus. Prosodia on jo jonkin aikaa ollut yksi eniten uutta tutkimusta tuottavista fonetiikan osa-alueista.

Nykyaikaisen tieteen kehyksessä fonetiikan kehityskulun voi sanoa alkaneen toden teolla 1700-luvulla. Tuolloin ei vielä ollut sellaisia tutkimusvälineitä, joilla ääntä ja sen eri ominaisuuksia olisi voinut täysin objektiivisesti kuvata, joten foneettinen kuvaus tahtoi jäädä kuulovaikutelmien ja niiden impressionistisen luonnehdinnan varaan. Tämä ei sinänsä ole niin suuri rajoitus havainnoille kuin saattaisi kuvitella, koska tarkkakorvainen – käytännössä usein musikaalisesti lahjakas – kuulija voi tehdä puheesta varsin hienovaraisiakin havaintoja. Ongelma on ennen kaikkea metodologinen: miten siirtää korvin havaittu tutkimustieto luotettavasti toisille tutkijoille? Tällainen havainnointi ei myöskään voi koskea artikulaatiota. Pelkän kinesteettisen, puhe-elimistön kulloistakin muotoa ja tilaa koskevan aistin varassa sitä on osoittautunut äärimmäisen vaikeaksi tavoittaa, eivätkä ihmisen ääntä koskevat havainnot koskaan ole suorassa suhteessa äänen akustisiin ominaisuuksiin. Tällainen tutkimustapa voi siis myös vääristää akustiikasta tehtyjä havaintoja.

Merkittävä edistysaskel oli keksiä keinoja visualisoida puhetta eri tavoin eli muuttaa se nähtäväksi, jolloin sitä saattoi myös luotettavammin mitata. Varhaisista tutkimusvälineistä kymografi keksittiin 1840-luvulla alun perin verenpaineen mittaamiseen, mutta sitä sovellettiin pian myös puheeseen (Suomessa esim. Äimä 1918a; 1918b). Koehenkilö puhui suukappaleeseen, painevaihtelut liikuttivat rummun kalvoa ja lopulta liikkeet piirtyivät nokipaperiin tasaisesti pyörivällä sylinterillä. Näin syntyneistä puhekäyristä saattoi jo mitata vaikkapa kestoja tai analysoida äänisignaalia. Laite oli yleisessä käytössä pitkälle 1900-luvun alkupuolelle.

Toinen merkittävä edistysaskel kokeellisen fonetiikan muotoutumisessa nykyiselleen oli äänentallennuksen keksiminen 1800-luvun lopulla. Vasta kun ääntä saattoi tallentaa, samaa puhejaksoa oli mahdollista

kuunnella useita kertoja peräkkäin ilman, että se olisi puhujan tuottamuksessa muuttunut. Koska kaikessa empiirisessä tutkimuksessa toistettavuus on yksi keskeinen kysymys, äänentallennus takasi mahdollisuuden tehdä fonetiikasta entistä empiirisempää.

Kolmannen edistysaskelen muodostavat ryhmänä kaikki ne tutkimusvälineet, jotka 1800-luvun lopusta alkaen ovat mahdollistaneet yhä uusien puhetapahtuman osatekijöiden empiirisen tarkastelun, esimerkiksi artikulaation eri kuvantamismenetelmät (kuten röntgen; ks. luku 3.1.) sekä lukuisat uudet, etenkin tietokoneaikana kehittyneet äänen akustiikan visualisoinnin ja käsittelyn tavat (ks. luku 3.2.). Kaikilla osa-alueilla tuotekehitys jatkuu vilkkaana, vaikkei sitä ensi sijassa ohjaakaan fonetiikan tutkimus vaan yleensä ulkoiset soveltavat syyt, kuten puheentunnistuksen tai tietokoneiden puhekäyttöliittymien kehitys.

## 2.2. Filosofiaa ja kielikäsitys

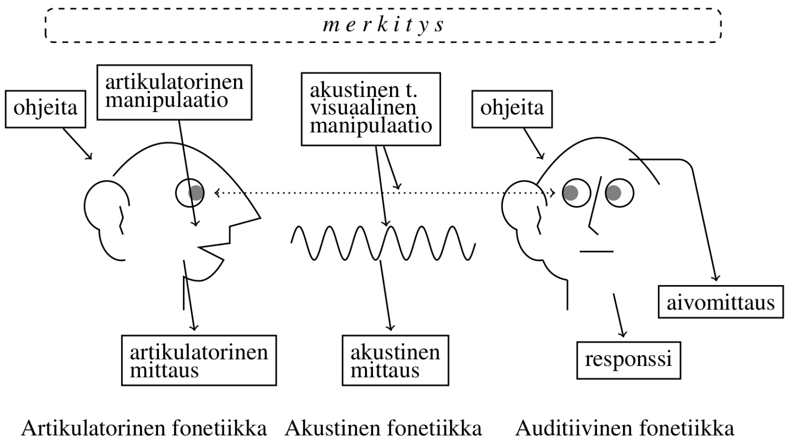
Foneettisia kuvauksia luonnehtivat jatkumot. Ihminen ei tarkalleen ottaen koskaan ”sano” esimerkiksi lyhyttä *a*- tai *ä*-vokaalia, vaan hän tekee tiettyjä asioita puhe-elimistöllään, joka jonkin ajan on jatkuvassa liikkeessä ja tuottaa toisen ihmisen kuultavissa olevaa, jatkuvasti muuttuvaa ääntä. Puhe-elimistö ei ole kuin auton vaihteisto, jossa on määräasemia määräasetuksille, sillä ei ole ennalta asetettuja tarkkoja ajallisia kestoja eri mittaisille puhesegmenteille, ja mikä tärkeintä, se ei juurikaan säilytä jotakin määrätilaa muuttumattomana pidempiä aikoja vaan on kaiken aikaa dynaamisesti liikkeessä.[1] Havaintokohteet, kuten ”äänteet”, ovat vain havainnon tuottama kuva tästä jatkuvasta saumattomasta muutoksesta. Todellisuudessa puheelle leimallista on jatkuva muutos.

Puheen toinen tunnusomainen piirre on koartikulaatio, ”myötä-ääntyminen”. Puheen eri osat, kuten artikulatoriset liikkeet eli puhe-eleet, eivät esiinny ajallisesti yksi toisensa jälkeen kuin rakennuspalikat vaan suuressa määrin toistensa kanssa ajallisesti limittäin. Tässä puhe eroaa kirjoituksesta, jossa näyttää olevan vain lineaarisesti peräkkäisiä toisistaan hyvin erotettavia yksiköitä, kirjaimia. Vaikka on totta, että yhden puhujan ääntämisen tuloksena syntyy äänisignaali eli ilmanpaine vaihtelee

lineaarisesti määrääjan, saman voi sanoa esimerkiksi orkesterin soittamasta musiikista. Silti siitä ei seuraa, että sinfonia olisi ajallisesti yksinomaan lineaarinen tapahtumasarja. Joskus onkin sanottu, että puhe on pikemmin sinfonia kuin yhden instrumentin sooloesitys, ja puhujan artikulatorista "suunnitelmaa" onkin kutsuttu partituuriksi.[2]

Kaiken lisäksi tiedetään, että puheen osien keskinäiset ajalliset suhteet muuttuvat joka tapauksessa kuulohavainnossa mm. siksi, että korkeammat taajuudet havaitaan nopeammin kuin matalammat ja puheen eri osien prosessointi aivoissa kestää muutenkin eri aikoja. Kun edelleen otetaan huomioon puheviestinnän kaikki osallistujat, on selvää, etteivät puheenvuorot ole yksiselitteisesti peräkkäisiä vaan usein päällekkäisiä. Ääritapauksessa voi olla hyvin monta puhujaa tuottamassa puhetta samanaikaisesti, jolloin ei voi edes puhua puheenvuoroista (esim. mielenosoituksissa).

Siinä missä (muu) kielitiede on pyrkinyt aina diskreetteihin, kategorisesti toisistaan erottuviin havaintokohteisiin (esim. suomessa *a* on eri asia kuin *ä* tai sanat *ratti* ja *rätti* merkitsevät eri asioita), fonetiikka hyväksyy kaikkialla havaintokohteissaan dynaamisen jatkuvuuden, jota merkityksen välittäminen jollakin tavoin ohjaa (ks. myös kuvio 1).




Kuvio 1. Puheketju ja tutkijan mahdollisuudet manipuloida ja havainnoida sitä.

# 3. Fonetiikan menetelmiä

Perinteisesti on ollut tapana jakaa fonetiikka kolmeen osa-alueeseen ns. puheketjun eri vaiheiden mukaan. Artikulatorinen fonetiikka keskittyy puheen tuottamiseen, akustinen fonetiikka itse puheääneen ja auditiivinen tai perseptuaalinen fonetiikka puheen havaitsemiseen. Yhtä lailla tutkimusmenetelmät voi jakaa karkeasti kolmia sen mukaan, mitä puheketjun vaihetta tutkija havainnoi tai mittaa, eli artikulatoriseen, akustiseen ja perseptuaaliseen tutkimukseen.

**Fonetiikka ja viittomakieli**

Viittomakielessä on vastaavat puheketjun vaiheet, vaikka perinteinen terminologia (mukaan lukien sana *fonetiikka* < kr. *fōnē* 'ääni') viittaa puhuttuun kieleen. Viittomakielen tutkimuksessa on joskus puhuttu esim. optisesta fonetiikasta akustisen sijaan.

Toisinaan tutkija vain havainnoi puhetta ja puhetilanteita eikä pyri tekemään järjestelmällisesti kontrolloituja kokeita. Tällöin kiinnostuksen kohteena on havaittujen aspektien eli havaintomuuttujien keskinäiset riippuvuussuhteet. Tiukemmassa mielessä *kokeellinen*-sanaa on tulkittu niin, että tutkija itse järjestelmällisesti manipuloi joitakin tutkittavan ilmiön piirteitä, ns. kontrolloituja tai kontrollimuuttujia, ja havainnoi toisia piirteitä selvittääkseen näiden mahdollisia keskinäisiä vaikutussuhteita.[3] Kun tutkija kontrolloi jotakin ominaisuutta, hän pyrkii varmistamaan, ettei ominaisuus riipu – ainakaan sillä kertaa – muista havaintoseikoista. Fonetiikan tutkimusmenetelmiä voi siis luokitella myös sen mukaan, manipuloiko tutkija puhetapahtumia järjestelmällisesti vai ei, ja jos manipuloi, missä kohtaa puheketjua tutkija niiden kulkuun vaikuttaa mahdolliset seuraukset nähdäkseen.

Tutkimuksen ajassa mahdollinen manipulointi aina edeltää havainnointia. Siitä ei silti seuraa, että manipulaatio puheketjukaaviossa aina

sijaitsisi havainnoinnin vasemmalla puolen, koska kuvion 1 hahmottelemasta puheketjumallista poiketen todellinen puhevirta ei ole lineaarinen prosessi, jossa on selvä alku ja loppu, vaan se on (vähintään) syklinen. Periaatteessa minkä tahansa vaiheen ilmiö voi vaikuttaa mihin tahansa tulevaan vaiheeseen.

Fonetiikka tutkii äänen käyttöä kielellisessä viestinnässä. Toisella tavoin ilmaisten kohteena on äänen ja merkityksen suhde.[4] Foneettisessakin tutkimuksessa olennainen muuttuja on näin ollen jollakin tapaa operationaalistettu merkitys. Toisaalta fonetiikassa merkitystä tyydytään käsittelemään hyvin epämääräisesti eikä pyritä analysoimaan merkitysrakenteita tai merkitysten kompleksisia suhteita. Yleensä riittää, että tutkija pystyy arvioimaan, onko kahden annetun puhunnoksen merkitys sama (tai samankaltainen) vai ei.

Merkityseron toimiessa kontrolloituna muuttujana koehenkilöitä voidaan ohjeistaa käyttäen kielen omaa oikeinkirjoitusta (ortografiaa), jos halutut merkityserot selviävät sen avulla. Joskus tutkijan täytyy olla kekseliäämpi. Esimerkiksi jos haluaa tutkia, onko suomen kielen ääntämyksessä eroa merkitysten ’kurkku (ruumiinosa)’ ja ’kurkku (vihannes)’ välillä, tilanne on hankalampi, koska merkitys toimii havaintomuuttujana. Jollakin tarkkuudella tutkija voi yrittää arvioida, mitä puhuja tarkoitti ja miten kuulija kuulemansa tulkitsi, mutta on selvää, että tällainen arviointi on subjektiivista ja altista virheille. Kummassakin tapauksessa on käytännössä usein mahdollista korvata ”merkitykset” sanoilla tai muilla oletetuilla kielellisillä yksiköillä.

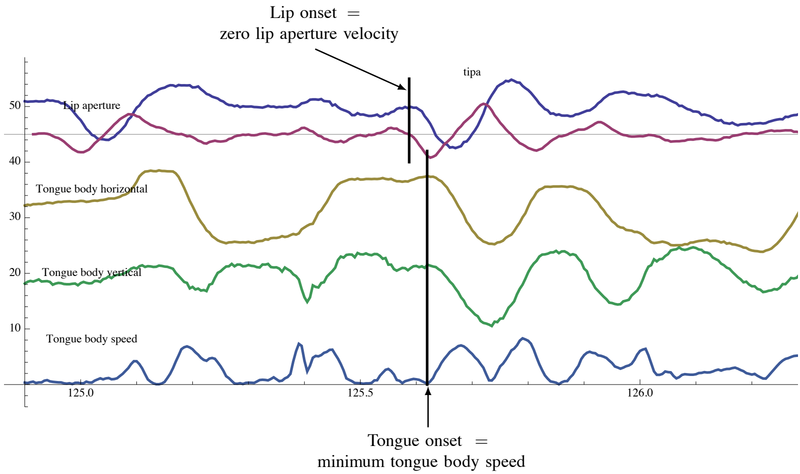
## 3.1. Artikulatorinen tutkimus

Fonetiikan vanhin ja luultavasti useimmille alan ulkopuolisille yhä tutuin tarkastelukulma on artikulatorinen. Siinä huomio kiinnittyy siihen, mitä puhujan puhe-elimistössä puhetta tuotettaessa tapahtuu. *Artikulaatio* suomentuu periaatteessa ’ääntämiseksi’, mistä helposti seuraa väärinkäsitys, että artikulatorinen fonetiikka liittyisi jotenkin ”oikeaan” tai ”kunnolliseen” ääntämykseen kuten puhetaidon oppaissa. Ei-preskriptiivisenä tieteenä fonetiikkaa kiinnostaa kuitenkin kaikki

ääntäminen samoin kuin ääntämiseen liittyvät muut fysiologiset tapahtumat, kuten hengittäminen ja lihaskontrolli.

Artikulaation empiiriseen tutkimukseen on ollut historiallisesti helpointa tarttua, koska jokaisella kielenpuhujalla on hallussaan ja varsin pitkälle hallinnassaan se elimistö, jossa puhe syntyy. Ihminen on myös sangen etevä matkimaan omassa puhe-elimistössään toisen ihmisen tuottamaa puhetta. Artikulaatiota voi kuitenkin vain rajallisesti tutkia matkimalla kuultua puhetta ja yrittämällä kinesteettisen tuntuman perusteella arvioida puhe-elimistön muotoa. Ensinnäkin tuntuma voi valehdella – minkä säilyneet varhaiset artikulatoriset kuvaukset armotta nykytutkijalle paljastavat –, ja toisekseen jo tieteellisen tutkimuksen luonne tapaa vaatia tutkijaa olemaan liikaa luottamatta itseensä havaintokohteena. Toisen henkilön puhe-elimistön havainnoiminen edellyttää välineitä tai vähintäänkin tutkimusolosuhteita, jotka tunkeutuvat syvälle tutkimuskohteena toimivan ihmisen normaaliin toimintaan, eivätkä voi jollakin lailla olla vaikuttamatta siihen. Tutkimusvälineet tahtovat aina jonkin verran haitata puhetta.

**Tyypillisen artikulatorisen tutkimuksen rakenne**. Artikulatorisessa tutkimuksessa ääntämisen seuranta tyypillisesti edellyttää koehenkilöjen tulevan (pysyvään tai väliaikaisesti rakennettuun) laboratorioon, jotta eri mittauslaitteet olisivat käytettävissä. Monet menetelmät ovat koehenkilöille verraten raskaita, eikä niitä voi käyttää kovin pitkään kerralla. Havainnointiaika on näin ollen rajallinen, joten ei ole mielekästä jäädä odottamaan, että tutkijaa kiinnostavia puhetapahtumia tapahtuisi spontaanisti. Tyypillisesti koehenkilöitä pyydetäänkin lukemaan ennalta määrättyjä tekstejä, esimerkiksi sanalistoja, lauseita tai kokonaisia tekstejä.




Kuvio 2. Esimerkki artikulaatiomittauksista: huuliaukko, kielen selkä (sijainti pysty- ja vaakasuunnassa sekä selän nopeus) (O'Dell, Šimko, Nieminen, Vainio, Lehtinen 2011).

Puhetta voi yrittää maanitella luonnollisemmaksi niin, että pyytää koehenkilöä painamaan annetun tekstin ensin mieleensä ja sitten ääntämään sen ilman kirjallista tekstiä. Tyypillisesti mittauslaitteisto synnyttää jatkuvasti muuttuvaa signaalia, esimerkiksi seuraa koehenkilön jatkuvia artikulatorisia liikkeitä. Koehenkilön puhe on hyvä samanaikaisesti äänittää siten, että äänen ja artikulaatiotietojen tahdistus eli synkronointi on jälkeenpäin mahdollista. Koehenkilön puheesta on myös yleensä helpompaa löytää tutkimuksen kannalta kiinnostavia vaiheita äänitteen avulla. Lisäksi äänitteen avulla voi tarkistaa, onko koehenkilö todella tuottanut, mitä pyydettiin.

Koska eri puhujat ovat joka tapauksessa anatomisesti erilaisia, artikulatoriset mittaukset eivät suoraan ole vertailukelpoisia yksilöstä toiseen. Eri ääntämyksiä on silti mahdollista vertailla ottamalla yksilön anatomia huomioon. Voidaan esimerkiksi todeta, äännetäänkö *i*-vokaali järjestelmällisesti eri tavoin kahdessa eri kontekstissa.

**Artikulatorisia mittauksia**. Yksinkertaisimpia artikulaation havainnoinnin tapoja ovat valo- ja videokuvaus sekä tutkijan oma (kinesteettinen eli liikeaistituntumaan perustuva) havainnointi. Koska artikulaatio

tapahtuu oleellisesti katseilta piilossa, eri kudosten alla, kaksi ensimmäistä tapaa tuottavat luonnollisesti rajallisesti tutkimustietoa. Tutkijan omaan tuntumaan perustuvan havainnoinnin rajoittuneisuus on puolestaan paljastunut jo fonetiikan historiassa, puhumattakaan siitä, että tutkijan asettuminen omaksi tutkimuskohteekseen on aina metodologisesti epätyydyttävä ratkaisu.

Röntgenkuvaus oli yksi ensimmäisiä merkittäviä tekniikoita paljastaa artikulaation katseelta piiloon jäävät tapahtumat. Röntgeniin liittyy kuitenkin suuria ongelmia. Ilmeisin niistä on nykyään hyvin tiedossa: röntgensäteily on suurina tai pitkittyneinä annoksina tutkittavan terveydelle haitallista, mistä syystä röntgenin käyttö välineenä on merkittävästi vähentynyt koko toisen maailmansodan jälkeisen ajan. Silti sen voi edelleen katsoa antavan parhaan ja tarkimman kuvan artikulaatioelimistä artikulaation aikana.

Terveyshaitat eivät ole kuvantamismenetelmän ainoa ongelma. Ensinnäkin puhe-elimistö koostuu suurelta osin pehmytkudoksesta, joka ei näy röntgenissä niin hyvin kuin kovat kappaleet kuten luut ja hampaat. Mitä pienemmistä pehmytkudososista on kyse, sitä huonommin ne röntgenissä näkyvät; tärkeimmän artikulaatioelimen kielen saa suurena lihaksena kuvassa kuitenkin näkyviin onneksi varsin hyvin. Toinen ongelma on se, että röntgenkuva on kaksiulotteinen projektio kolmiulotteisesta artikulaatioavaruudesta eli litistetty kuva, jossa ei voi yhtä aikaa saada näkymään kaikkia ulottuvuuksia. Usein käytetään vain sivu- eli profiilikuvakulmaa. Tarkastelua voisi helpottaa, jos olisi mahdollisuus samanaikaisiin röntgenkuviin useammasta kuvakulmasta, mutta tämä lisäisi säteilyannostusta entisestäänkin ja on siten yleensä mahdoton toteuttaa. Lisävaikeuden muodostaa alaleuan luun suuri, näkyvyyttä haittaava massa. Toisaalta röntgenkuvaus on mahdollista myös videona, jolloin aikaulottuvuus saadaan mukaan.

Röntgenin sijaan voi käyttää muita läpäiseviä menetelmiä, esimerkiksi ultraääntä. Ultraäänellä ei ole samanlaisia haittoja kuin röntgensäteilyllä, mutta sen antama kuva on vastaavasti epätarkempi ja vaatii tarkempaa tulkintaa. Kuten röntgeniä myös ultraääntä voi käyttää joko staattisina kuvina tai videona, mutta röntgeniin verrattuna ultraäänivideo voi huomattavasti helpottaa tulkintaa: mikä yksittäisessä

tilannekuvassa jää epäselväksi, saattaa elävästä kuvasta käydä paremmin ilmi. Ajallinen erotustarkkuus röntgenissä ja ultraäänessä on samaa luokkaa eli kummallakin pystyy tarkastelemaan myös varsin lyhytkestoisia tapahtumia.

Yleisesti ottaen kaikkien läpäisevien kuvausmenetelmien ongelmana on kiintopisteiden asettaminen tarkasteltaville puhe-elimille. Esimerkiksi kielen nousu voi johtua paitsi siitä, että puhuja nostaa kieltä, myös siitä, että puhuja muuttaa päänsä asentoa niin, että liikkumattomana suussa pysynyt kieli nousee. Koska koko puhe-elimistö – ainakin sen supraglottaalinen osa eli kaikki kurkunpäästä ylöspäin – sijaitsee hyvin liikkuvassa osassa ihmisvartaloa, kiintopisteiden asettaminen voi olla hankalaa ja edellyttää joko kuvantamislaitteiston kiinnittämistä tiukasti määräasentoon suhteessa päähän (ks. kuva 1) tai pään



Kuva 1. Ultraäänilaitteisto kiinnitystelineineen. Signaali siirtyy analyysia varten tietokoneelle.

kiinnittämistä tiukasti mittauslaitteistoon; kumpikin tapa tekee informantin olosta enemmän tai vähemmän epäluonnollisen, mikä voi heijastua puheeseen.

Elektromagneettinen artikulografia (EMA) ratkaisee ongelman siten, että samanaikaisesti seurataan pään liikkeitä ylimääräisillä sensoreilla, jotta mittausten jälkeen voidaan laskea artikulaattoreiden liikkeitä suhteessa pään liikkeisiin. Tässä mittaustekniikassa kiinnitetään sensoreita väliaikaisella liimalla tutkittaviin kohtiin, esim. huuliin, kielen kärkeen, laiteisiin, lapaan ja selkään, ylä- ja alahampaisiin tai mahdollisesti kitapurjeeseen. Laitteisto pystyy seuraamaan sensorien liikkeitä, kunhan sensorit pysyvat noin kuutiometrin kokoisen kehyksen tai häkin sisällä, jonka kulmissa on voimakkaat kestomagneetit.[5] Tyypillisesti sensorit kiinnitetään myös korvien taakse ja nenän yläpuolelle pään liikkeiden seurantaa varten. Artikulografia välttää ultraäänikuvantamiselle tyypillisen ongelman pään fyysisestä kiinnittämisestä johonkin ulkoiseen kehykseen, mikä vähentää tutkimustavan invasiivisuutta, mutta toisaalta tutkimushenkilö on sijoitettava suureen häkkiin ja hänen suuhunsa on kiinnitettävä sensorit johtoineen, jotka ohuudestaan huolimatta suusta ulos pistävinä ja huulta vähäisellä mutta havaittavalla massallaan painavina eivät voi olla jonkin verran vaikuttamatta puhujan artikulaatioon. EMA-laitteisto on Suomessa käytössä ainoastaan Helsingin yliopistossa.

Ei-läpäiseviä artikulatorisia tutkimusmenetelmiä ovat palatografia ja labiografia. Niihin ei sisälly suoranaisia terveysriskejä, mutta ne myös rajoittuvat tuottamassaan tutkimustiedossa yhteen ääniväylän osaan: palatografia kitalaen ja kielen kosketukseen, labiografia huuliin. Vanhastaan palatografiasta puuttui täysin aikaulottuvuus eikä se myöskään sallinut tutkia pidempiä puhejaksoja, mutta nykyisin dynaamisella eli elektropalatografialla voi kerätä jatkuvaa ajallista tietoa siitä, missä kieli koskettaa kitalakea. Myös palatografia on niin perinteisessä kuin nykymuodossaankin varsin invasiivinen tekniikka, joka ei salli luonnollisia puhetilanteita ja vaatii koehenkilöltä yhteistyöhalukkuutta.

Ei-läpäisevänä voinee pitää myös elektromyografiaa (EMG) eli lihasten sähköisen toiminnan havainnointia. Puheentutkimuksen tekniikkana se on jo vanha ja vakiintunut. Menetelmä tuottaa tietoa artikulaatiolihasten toiminnasta, mistä edelleen voi epäsuorasti päätellä seikkoja

myös artikulaatioelinten kulloisestakin asemasta. Sen erityisansiona voi pitää sitä, että se tuottaa tietoa puhe-eleestä sen luontaisilla yksiköillä eli se astuu askelen taaksepäin puheketjussa muihin mainittuihin kuvantamismenetelmiin verrattuna suomalla näkymän siihen, mitä elimistössä tapahtuu juuri ennen havaittavaa artikulaatiota.

Yksi hankalimmista artikulatorisista tutkimuskohteista on kurkunpää sointia tuottavine äänihuulineen. Tutkittavaksi on vaikea saada esimerkiksi puhdasta sointiääntä eli äänen sointia sellaisena kuin se kurkunpäästä lähtee ilman, että kurkunpään yläpuolinen ääniväylä resonanssiominaisuuksineen vaikuttaisi siihen. Vaikka mikrofoni kiinnitettäisiin kaulaan kurkunpään kohdalle, resonansseilla voi silti olla merkittävästi häiritsevä vaikutus. Yhtenä ratkaisuna on ollut käyttää kuolleen ihmisen kurkunpäätä, jonka voi irrottaa tutkittavaksi erikseen tuottamalla ilmavirran sen läpi. Tällöin kuitenkin saadaan selville lähinnä kurkunpää-äänen fyysiset rajat, ei sen dynaamisia ominaisuuksia eikä ennen kaikkea sitä lihaskontrollia, jolla elävä ihminen kurkunpää-ääntään puheen osana käyttää. Edelleen äänihuulten värähtely on niin nopeaa – taajuus jopa satoja kertoja sekunnissa –, ettei kaikkien kuvantamismenetelmien aikaerotuskyky ole vallankaan aiemmin riittänyt sen tallentamiseen. Nykyisin suurinopeusvideokuva on kuitenkin jo arkipäivää, joten ongelmaksi jää lähinnä saada kamera kuvaamaan kaulan lihasten ja muiden kudosten sekä varsin tukevien rustorakenteiden taakse piiloutunutta tapahtumaa. Käytettyjä tekniikoita ovat mm. röntgenkuvantaminen (edellä mainittuine ongelmineen), koehenkilön kannalta varsin invasiivinen tähystyskamera tai 1950-luvulla kehitetty elektroglottografia, jossa johdetaan suuritaajuuksinen sähkövirta kurkunpään lävitse ja mitataan muutokset sen impedanssissa. Kaikissa tutkimusmenetelmissä on puutteensa.

Muista menetelmistä on mainittava ilmanpaineen ja -virtauksen mittaus. Kumpikin mittaus on periaatteessa mahdollista tehdä täysin ulkoisesti, mutta käytännössä ne kuitenkin tapaavat edellyttää varsin ankaraa invasiivisuutta. Esimerkiksi suu- ja nenäaukon peittävä (oronasaalinen) ns. Rothenberg-maski on yksi tällainen apuväline. Menetelmän ajallinen erottelukyky ei kuitenkaan riitä yksittäisten äänteiden tarkkailuun.

**Artikulatorisia manipulointimahdollisuuksia**. Olisi ihanteellista, jos puhuja pystyisi tarkoin ja pienin askelin varioimaan omia artikulatorisia

liikkeitään ja lihastensa käyttöä, jotta voitaisiin tarkkailla seurauksia akustiikassa tai havainnossa. Puhujalla ei ole kuitenkaan tällaisia kykyjä. Artikulaatio muuttuu jopa silloin, kun puhuja tietoisesti pyrkii ääntämään saman puhunnoksen täsmälleen samalla tavalla.

Luultavasti nykyään tavallisin tapa kiertää tämä este on korvata oikea puhuja artikulatorisella synteesillä. Siinä puhetta tai tarkemmin puheenkaltaista ääntä tuotetaan tietokoneohjelmalla, joka mahdollisimman tarkasti pyrkii mallintamaan ihmisen puheentuottamismekanismia. Etuna oikeaan ihmiseen verrattuna on se, että mallia voi säätää luotettavasti ja pikkutarkasti esimerkiksi lisäämällä tietyn lihaksen supistusvoimaa, ja ohjelman synnyttämää akustista tulosta voi tarkkailla sellaisenaan tai esimerkiksi soittaa (ihmis)kuulijoille havaintokokeessa. Mallia voi jopa säätää tavalla, joka olisi ihmiselle mahdotonta, esimerkiksi suurentamalla ja pienentämällä eri anatomisia osia, jotta nähtäisiin, miten ihmisten luonnolliset anatomiset erot vaikuttavat ääneen. Haittapuolena on tietenkin se, ettei malli kuitenkaan koskaan ole todellinen ihminen eikä kaikkia olennaisia artikulatorisia seikkoja ole välttämättä kyetty ottamaan huomioon. Artikulatorisen synteesin kehittäminen edellyttääkin edeltävää tietoa nimenomaan oikean ihmisen artikulaatiosta.

Oikean puhujan manipulointiin on monia mahdollisuuksia, esimerkiksi purentakappale (engl. *bite-block*; esim. Fowler & Turvey 1980), puhuminen oudolla tavalla (esim. suu tiukasti kiinni), puheen toistaminen vain toistuvaa tavua käyttäen (engl. *reiterant speech*; esim. Kelso, Vatikiotis-Bateson, Saltzman & Kay 1985) jne. Ihmisellä on varsin hämmästyttävä kyky tuottaa ymmärrettävää ja jopa varsin oikealta kuulostavaa puhetta myös eri tavoin häirittynä, mikä osoittaa oivasti, kuinka puhujalla on hallittavissa oleva malli paitsi omasta ääntöelimistöstään myös sen ja havaintovasteen suhteesta.

Samoin on osoitettavissa, että puhuja käyttää eri palautekanavia puheentuottonsa monitorointiin. Jos palautejärjestelmää häiritään, puhuja tyypillisesti ”korjaa” artikulaatiotaan, vaikka korjaus lopullisen puhetuotoksen kannalta saattaa häirinnän vuoksi nimenomaan osua vikaan. Keinoja ovat esimerkiksi väliaikainen puuduttaminen, kohinan toistaminen puheen aikana sekä viivästetty äänipalaute (engl. *delayed audio feedback* eli *DAF*), jossa puhujan korviin soitetaan hänen omaa ääntään

jonkin verran viivästettynä, jolloin puhujan oma lihastuntuma ja hänen havaintonsa omasta äänestään ajautuvat ristiriitaan.

## 3.2. Akustinen tutkimus

Akustinen tutkimus tutkii itse (puhe)ääntä fysikaalisena ilmiönä, jolloin koehenkilön ei enää tutkimusdataa kerättäessä ole välttämätöntä olla läsnä. Itse äänidata on tietenkin silti kerättävä koehenkilöiltä, ja koska foneettinen tutkimus luonteeltaan tarkastelee hienovaraisia jatkumoita ja moninaisten piirteiden keskinäistä dynamiikkaa, käytettäville äänitteille asetetaan usein – vaikkei välttämättä aina – tiukemmat laatuehdot kuin suuressa osassa muuta kielentutkimusta. Ei riitä, että tutkija saa sanoista selvää, vaan hänen on voitava havainnoida niitä ilmiöitä, jotka mahdollistavat puheen, tai yhtä hyvin, jotka vaikeuttavat sitä.

Äänityksen laatuun vaikuttavat monet tekijät. Osa on peräisin laitteistosta, jolloin tutkija voi vaikuttaa niihin ainoastaan valitsemalla parempilaatuisia laitteita. Tällainen ominaisuus on esimerkiksi laitteen taajuusvaste eli se, kuinka luotettavasti äänite tallentaa alkuperäisen äänen kaikkia taajuusalueita suhteessa toisiinsa. Osa häiriötekijöistä taas aiheutuu äänitysympäristöstä ja laitteiden käytöstä. Niihin voi äänitystilanteessa mahdollisesti vaikuttaa. Näihin kuuluvat esimerkiksi ulkoisten häiriöäänten määrä ja signaali-kohinasuhde. Lisäksi koska ääntä käsitellään nykyisin yleensä tietokoneella, on otettava huomioon tekijät, jotka aiheutuvat luonteeltaan analogisen (jatkumonluonteisen) äänisignaalin digitoimisesta tietokoneen käsiteltävissä olevaan muotoon. Edelleen on huomattava, että itse digitoijan[6] laatu voi myös vaikuttaa siihen, kuinka luotettavasti digitaalinen äänidata vastaa alkuperäistä analogista signaalia. Moniin tekijöistä voi olla hyvin vaikea vaikuttaa. Laitevalmistajien lupauksiin ei kannata sokeasti luottaa, vaan on tarkistettava, mitä esimerkiksi lehtien testit laitteiden laadusta ovat paljastaneet.

**Laatutekijät äänen digitoinnissa**

Äänen digitoimisessa on helppo tehdä karkeita virheitä. Seuraavat neljä tekijää kannattaa tarkistaa yleisten laatutekijöiden ohella.

**Näytteenottotaajuus**. Kuinka usein äänisignaalista otetaan näyte. Musiikki-CD-levyjen näytteenottotaajuus on 44,1 kHz (44 100 kertaa sekunnissa), joka riittää kuvaamaan taajuusalueen 0–22,05 kHz. Suurin osa puheen informaatiosta löytyy taajuudelta 0–10 kHz, joten tämä standardiksi muodostunut taajuus hyvin riittää puheen tutkimiseen.

**Bittisyys**. Kuinka suurta binaarilukua käytetään kuhunkin näytteeseen. 8 bittiä riittää 256:een eri tasoon, 16 bittiä 65 536:een jne. Mitä vähemmän tasoja on käytettävissä, sitä pienempi on näytteen hiljaisimman ja voimakkaimman äänen välinen ero. 16-bittinen äänitys on riittävä puheen tutkimiseen.

**Pakkaus**. Monet äänentallennuslaitteet pakkaavat ääntä automaattisesti. Pakkaaminen vähentää äänisignaalin dataa niin, että auditiivinen vaikutelma säilyy mahdollisimman samankaltaisena. Tutkimustarkoituksiin kerätyssä äänessä pakkaaminen aiheuttaa yleensä ennakoimattomia ongelmia, joten pakkausta on syytä välttää.

**Äänitystaso**. Etenkin digitaalisessa äänityksessä on tärkeää säätää äänitystasoa siten, ettei digitoinnin katto ylity (ainakaan liian usein), koska tällöin informaatiota katoaa. Toisaalta ei ole suositeltavaa käyttää joidenkin laitteiden tarjoamaa automaattista voimakkuuden säätöä muun muassa siksi, että silloin asetus vaihtelee ajan myötä eikä taso siten ole vertailukelpoinen näytteen sisällä.

Puhe-elimistössä syntyneen puheäänen tutkiminen muuten kuin impressionistisesti vaatii laitteita, joilla visualisoida ääntä ja mitata sen ominaispiirteitä. Akustiikan tutkiminen ei sinänsä vaadi puheen tuottaneen henkilön läsnäoloa; tässä mielessä akustinen tutkimus vapauttaa koehenkilön laboratoriosta. Toisaalta fonetiikan tarpeisiin sopiva akustinen aineisto on usein laatuvaatimusten takia äänitettävä olosuhteissa, jotka ovat tiukan laboratoriomaiset eivätkä siten suosi vapaata, luonnollista puhetta.

Vielä pitkälle yli 1900-luvun puolenvälin akustiikkaa visualisoivat järjestelmät olivat kalliita ja mutkikkaita erikoislaitteita, joita saattoi löytää lähinnä tutkimuslaitoksista. Henkilökohtaisten tietokoneiden kehityksen myötä tilanne muuttui nopeasti. Alkuun tosin tilanne muuttui vain sikäli, että vaikka äänen akustiseen tutkimukseen soveltuva laite – tietokone, johon voi syöttää ja josta voi saada ulos ääntä – oli helposti saatavilla, visualisoinnin tekevät ja mittaukset mahdollistavat ohjelmat olivat yhä kalliita ja harvojen ulottuvilla. Tämäkin tilanne helpottui oleellisesti 1990-luvulla, kun ilmaisesti verkossa jaettava Praat-ohjelma (Boersma & Weenink 1996; 2018) tuli saataville. Nykypäivänä puheketjun akustinen vaihe lienee helpoin vaihe tehdä havaintoja tai manipulaatioita. Vähimmäislaitteistoksi riittää tietokone ohjelmistoineen.

**Tyypillinen akustinen tutkimus**. Akustisesta tutkimuksesta esimerkkinä toimii vokaalien formanttimittaus. Akustisesti vokaalit eroavat toisistaan formanteiltaan, jotka ovat vokaalin äänen spektrissä verraten voimakkaita taajuusalueita, jotka aiheutuvat ääniväylän resonansseista kurkunpäässä syntyneen äänen kulkiessa sen läpi. Resonanssit riippuvat ääniväylän kulloisestakin muodosta, mm. kielen ja huulten asennosta, mikä selittää, miksi formantit ovat vokaalien kannalta niin olennainen akustinen piirre: ne paljastavat epäsuorasti ääniväylän muodon eli puhe-elinten konfiguraation. Varsinkin kaksi alinta formanttia (F1 ja F2) ja vähemmässä määrin seuraavakin (F3) ovat osoittautuneet tärkeiksi vokaalien erottelussa.

Tavoitteena tutkimuksessa voi olla selvittää esimerkiksi, paljonko jonkin kielen vokaalit eroavat toisistaan akustisesti. Eräs haittatekijä on puhujien muodoltaan eroavat ääniväylät, mistä seuraa, ettei formanttimittauksia voi suoraan verrata yksilöstä toiseen. Varsinkin ääniväylän kokonaispituus vaikuttaa formanttien sijaintiin, joten esimerkiksi naisilla ja miehillä on usein hyvin erilaisia formanttiarvoja samoissakin vokaaleissa. Yleistyksiä voi kuitenkin tehdä koehenkilöryhmässä turvautumalla normalisaatioon eli käyttämällä formanttilukemia suhteutettuna puhujan omaan formanttiavaruuteen. Tilastollisessa analyysissa ääniväylien arvioidun pituuden voi ottaa huomioon, kunhan puhujilta on mittauksia tarpeeksi monesta vokaalista.

Formanttien taajuuksia voi arvioida visuaalisesti suoraan juoksevan puheen spektrogrammista tai esimerkiksi vokaalin keskiosasta laskettusta yksittäisestä spektristä (ks. jäljempänä kohta Akustisia mittauksia). Lisäksi on joukko laskennallisia rutiineja, jotka pyrkivät arvioimaan formantit automaattisesti. Ne voivat olla hyvin täsmällisiä mutta tyypillisesti vaativat puhujakohtaista säätelyä. Tutkijan täytyy joka tapauksessa rajata puhesignaalia, jotta mittaus kohdistuisi juuri mitattavaan vokaaliin.

Usein tutkija antaa puhujalle joukon sanoja tai lauseita äännettäväksi, joten kontrolloituna muuttujana voi toimia mikä tahansa puheen oletettu rakenteellinen seikka: itse vokaalin identiteetti, vokaalin äänteellinen ympäristö, sanojen rakenne, prosodia jne. Tällöin on suotavaa, että äänitetään samoja tapauksia samaltakin puhujalta toistamiseen, jotta voidaan arvioida, kuinka paljon mittauksissa on satunnaista vaihtelua. Foneettisia ilmiöitä voi jossain määrin tutkia nykyään myös äänitetietokannoista, kuten Suomen kielen nauhoitearkistosta (ks. Haastattelu aineistonkeruun metodina sosiolingvistiikassa tässä kirjassa).

Toinen mahdollisuus on käyttää verraten spontaania äänitettyä puhetta, jolloin ei ole varsinaisesti kontrolloituja muuttujia. Se tarkoittaa, että esimerkiksi vokaalin identiteettikin on tutkijan arvioitava itse, ympäristöä ei voi kontrolloida eikä toistoja saada lainkaan. Silloin on tärkeää, että aineisto on suuri, jotta silti olisi mahdollista muodostaa luotettava kuva tutkittavasta ilmiöstä.

**Akustisia mittauksia**. Kestot ovat akustisen mittauksen kohteista luultavasti yksinkertaisimpia mitata mutta eivät välttämättä yksinkertaisimpia tulkita oikein. Ensinnäkin on muistettava, että puhe on fysikaalinen jatkumo, jonka saavat aikaan useiden rinnakkaisten ja toisiinsa nähden suhteellisen itsenäisten artikulaattorien liikkeet; puheessa on vain harvoja luontaisia katkoskohtia. Keston mittaus puolestaan kohdistuu johonkin segmenttiin – äänteeseen, tavuun, sanaan, lausumaan tms. Havaintojatkumo on siis ensin jollakin tavoin segmentoitava. Toisekseen kestoja mitataan yleensä tavallisina ajan yksiköinä, esimerkiksi milli- tai senttisekunteina, mutta tutkijan on otettava huomioon, että kielenpuhujan kannalta puheen kestot eivät suoraan kytkeydy tähän ”fysikaaliseen” aikaan. Akustisten segmenttien havaittu kesto on tyypillisesti suhteessa ympäröivien elementtien kestoihin kompleksisin tavoin. Kestot ovat silti

varsin suosiollinen tutkittava ilmiö, koska ne säilyvät havaintokelpoisina usein hyvinkin huonolaatuisessa signaalissa.

Ajallisella ulottuvuudella on myös toinen puoli, nopeus eli tempo. Usein ajallinen kesto niillä segmenteillä, joiden ominaiskesto on lyhyt, ymmärretään helpommin 'kestoksi', kun taas ominaiskestoltaan pitkien segmenttien kesto tulkitaan 'nopeudeksi'. Periaatteessa voitaisiin lyhyttä vokaalia yhtä hyvin kutsua nopeaksi vokaaliksi ja pitkää vastaavasti hitaaksi. Yleisemmin temposta kuitenkin puhutaan vasta, kun havaintoyksikkönä on vähintään tavun mittainen segmentti.

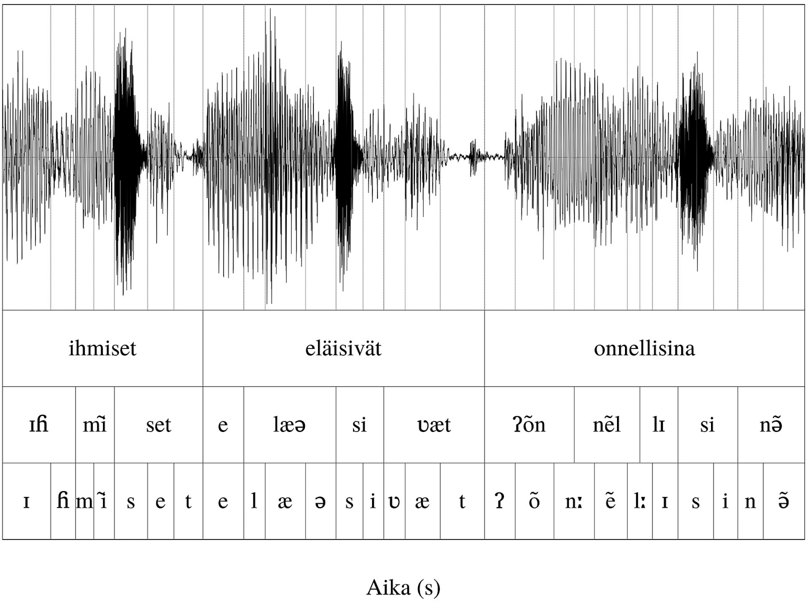
Äänenkorkeus on toinen piirre, joka säilyy signaalissa varsin hyvin. Tämä johtuu siitä, että äänenkorkeus kertautuu puhesignaalin muinakin piirteinä. Lähtökohtaisesti äänenkorkeuden aiheuttaa äänihuulissa syntyvä jaksollinen sointiääni, minkä vuoksi äänenkorkeus ilmenee vain puheen soinnillisissa jaksoissa. Soinnittomat konsonantit tai muu soinnittomuus, esimerkiksi kuiskaus, katkaisevat äänenkorkeuskäyrän, ja katkoksen jälkeen äänenkorkeus usein hetkellisesti muuttuu nopeasti. Käytännössä tästä seuraa, että havaintovaikutelmaa vastaavan katkottoman ja pehmeästi muuttuvan äänenkorkeuskäyrän aikaansaaminen ei mittauslaitteistoilla ole aivan yksinkertaista eikä kaikilta osin lainkaan mahdollista manipuloimatta käyrää käsin, paitsi jos varmistaa katkottomuuden etukäteen käyttämällä informanteilla äänitettävänä testilauseena *Laina lainaa lainalle lainan* -tyyppistä, yksinomaan soinnillisista äänteistä koostuvaa tekstiä.

Edelleen on huomattava, että äänenkorkeudessa on kyse taajuudesta eli frekvenssistä, mikä tarkoittaa jonkin tapahtumista niin ja niin monta kertaa jossakin ajallisessa yksikössä. Äänenkorkeus on näin sidoksissa aikaan. Yksinkertaisimmillaan tästä seuraa, että jos tarkasteltava ajanjakso on esimerkiksi 0,001 s (1 ms) ja tarkastelun kohteena 100 Hz:n äänenkorkeus (matalahko miesääni), havaintoikkunaan ei välttämättä mahdu yhtään tapahtumaa: 100 Hz:n taajuus tarkoittaa, että jotakin tapahtuu – äänihuulet avautuvat ja sulkeutuvat – sata kertaa sekunnissa eli yksittäinen jakso kestää 0,01 s (10 ms). Havaintoikkunan kohdalla ei välttämättä näy mitään.

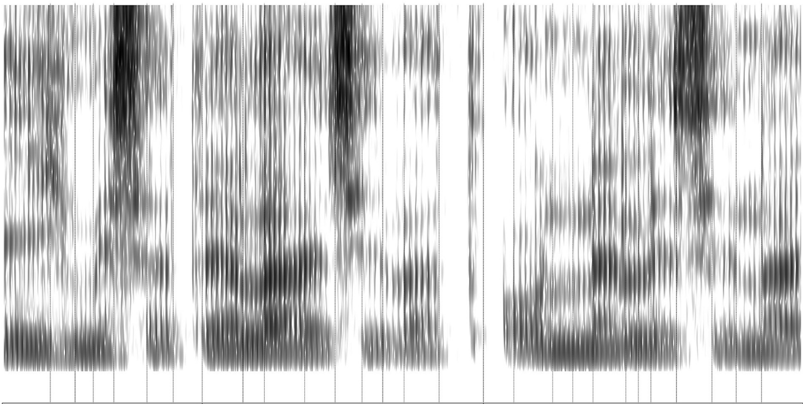
Äänenkorkeuskäyrän ohella voi seurata intensiteettikäyrää. Vaikka intensiteetti on yksi-yksisessä vastaavuussuhteessa havaittuun

äänenvoimakkuuteen, intensiteetin hetkittäisiä eli mikromuutoksia seurataan usein pikemmin muista syistä, esimerkiksi tutkimusvaiheessa segmentoinnin helpottamiseksi.

Toinen yleinen akustisten mittausten kohde on puheen spektri eli puheen taajuusominaisuudet. Ne ovat huomattavasti kestoja herkempiä äänitteen laadulle ja käsittelylle. Esimerkiksi äänen pakkausalgoritmit (kuten MP3) suunnitellaan yleensä niin, että ne säilyttävät mahdollisimman hyvin nimenomaan musiikin melodian ja havaintovaikutelman samalla, kun ne poistavat äänestä niin paljon tietoa kuin suinkin pakatakseen sen mahdollisimman pieneen tilaan. Puheentutkijan kannalta pakkausalgoritmit voivat poistaa oleellista tutkimustietoa.




Kuvio 3. Esimerkki ääniaaltokäyrästä ja sen litteraatista.



| | | |
|---|---|---|
| ihmiset | eläisivät | onnellisina |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ɪɦ | mĩ | set | e | læə | si | ʋæt | ʔõn | nẽl | lɪ | si | nə̃ |

| | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ɪ | ɦ | m | ĩ | s | e | t | e | l | æ | ə | s | i | ʋ | æ | t | ʔ | õ | nː | ẽ | lː | ɪ | s | i | n | ə̃ |

Kuvio 4. Esimerkki spektrogrammista ja sen litteraatista. Puhenäyte on sama kuin kuviossa 3.

Ääniaaltokäyrä on usein ensimmäiseksi käytetty väline puheääntä segmentoitaessa eli jaettaessa havainto- tai analyysiyksikköjä vastaaviin osiin (kuvioiden 3 ja 4 litteraattien kolme eri tasoa). Ääniaallolla on kaksi ulottuvuutta, x-akselilla aika ja y-akselilla äänen ilmanpaineen vaihteluja vastaava amplitudi, jonka voi mieltää ehkä helpoimmin ilmamolekyylin siirtymäksi lepotilastaan eli 0-tasolta. Segmentoinnin lisäksi ääniaaltokäyrä takaa siis jo mahdollisuuden ajan ja amplitudin mittaamiseen. Puheentutkimukselle tyypilliset kestomittaukset (äänteiden kesto, tavujen kesto, kestojen muutos ympäristön seurauksena jne.) on mahdollista tehdä ja usein tehdäänkin jo tässä näkymässä, muita näkymiä mahdollisesti vain segmentointiapuna käyttäen. Samoin on mahdollista mitata soinnillisen äänen taajuus yksinkertaisesti ääniaallon keston käänteislukuna ($f = 1/t$); soinnittoman äänen sijoittuminen eri taajuuksille vaatii jo spektrin laskemista.

Aina keksimisestään asti yksi keskeinen foneetikon työkalu on ollut edellä mainittu spektrogrammi (ks. kuvio 4).[7] Spektrogrammi yhdistää samaan analyysiin spektrin ja aikaulottuvuuden. Sitä voi pitää sarjana tietyin aikavälein otettuja äänen spektrejä, jotka nähdään ajallisesti x-akselille järjestettyinä, "ylhäältä päin" esitettyinä korkeuskäyrinä siten, että taajuusasteikko esitetään y-akselilla ja kunkin taajuuden amplitudi näkyy tummuutena (tai eri väreissä, käytetystä välineestä riippuen). Spektrogrammin ehdottomana hyvänä puolena on ollut nimenomaan näiden kolmen asian – ajan, taajuuden ja amplitudin – saaminen ymmärrettävästi ja selkeästi samaan kuvaan. Ääniaaltokäyrän ohella spektrogrammi onkin jo vuosikymmeniä ollut äänen akustisen tarkastelun perusväline. Käytännön artikulatorisessa työskentely-ympäristössä, kuten Praatin analyysiikkunassa, nämä kaksi esitystapaa yleensä tukevat toisiaan.

Myös spektrogrammi voi olla yksi segmentoinnin apuvälineistä, mutta sen varsinaisena vahvuutena on mahdollisuus tarkastella spektrin muutosta suhteessa aikaan. Puheentutkimuksessa yksi yleisesti käytetty äänen spektraalinen ominaisuus ovat jo mainitut formantit, jotka vastaavat muuttuvan ääniväylän kulloisiakin resonansseja eli ovat tulosta ääniväylän kurkunpää-äänelle aiheuttamasta suodatusfunktiosta. Ne näkyvät spektrogrammissa leveinä tummina vaakasuuntaisina vöinä tai raitoina, ja ne on tapana numeroida alhaalta ylöspäin ykkösestä alkaen: F1, F2, F3 jne. Esimerkiksi vokaaliavaruus on hyvin hahmotettavissa jo kolmen (tai jopa kahden) alimman formantin avulla, sillä kullakin vokaalilla nämä resonanssit sijaitsevat eri kohdissa. Sama informaatio löytyy periaatteessa myös spektrikuvasta, joka on ajallisesti lyhyempi, mutta käytännössä hetkittäisen vaihtelun poissulkemiseksi on usein helpompi käyttää ajallisesti laajempaa spektrogramminäkymää.

Spektrogrammeja on useanlaisia sille asetetuista parametreista riippuen. Parametreja ovat aikaikkunan pituus, analyysissa käytettävä aika- ja taajuusaskel sekä aikaikkunan muoto. Aikaikkuna tarkoittaa tässä sen äänikatkelman kestoa, josta spektri lasketaan; koska aika ja taajuus ovat erottamattomassa suhteessa toisiinsa (jo esillä olleen kaavan $f = 1/t$ mukaisesti), niitä ei käytännössä koskaan täysin voi erottaa toisistaan. Mitä tarkempi analyysi on ajan kannalta, sitä epätarkempi se on taajuuden kannalta ja toisin päin. Eri parametrein voi tutkia tarkemmin erilaisia

asioita. Kaksi perinteisesti käytettyä spektrogrammin muotoa ovat kuviossa 4 esitetty ajan suhteen tarkempi leveäkaistaspektrogrammi (engl. *wide-band spectrogram*), joka sopii erityisesti formanttien mittaukseen, ja ajan suhteen epätarkempi kapeakaistaspektrogrammi (engl. *narrow-band spectrogram*), jota voi käyttää esimerkiksi äänenkorkeuskäyrän (F0-käyrän eli kurkunpää-äänen perusvärähtelyn taajuuden[8]) analysointiin.

Akustisia manipulointimahdollisuuksia ovat muun muassa:

- (Uudelleen)syntetisointi. Käytetään todellista puhetta joko mallina tai lähteenä syntetisoidulle näytteelle, jolla tutkimusta varsinaisesti tehdään. Syntetisointi on erinomainen keino tutkia piirteitä, joita ei joko kohdata luonnollisessa puheessa lainkaan tai joiden tuotattaminen todellisilla puhujilla aiheuttaa ei-toivottuja seurannaisilmiöitä jossakin muualla. Esimerkkinä voisi olla äännekestojen tutkiminen. Syntetisoidussa puheessa on mahdollista tuottaa portaattomasti erilaisia äännekestoja, kun taas puhuja yleensä hallitsee luontevasti vain ne kestoasteet, jotka hänen puhumaansa kieleen kuuluvat.
- Kohinan lisäys, suodatukset. Muuttamalla äänisignaalia eri tavoin voidaan helpommin tutkia, mihin kuulijan havainto kohdistuu ja millä tavoin. Esimerkiksi ns. kaistanpäästösuodatuksen avulla, jossa äänestä pääsee läpi vain jokin taajuusalue (esim. 300–3 000 Hz), voidaan todeta, että kuulija havaitsee puheen perussävelen silloinkin, kun sen taajuus ei edes sijaitse suodatuksesta läpi päässeellä alueella.

## 3.3. Perseptuaalinen tutkimus

Puheen havaitseminen ja vastaanottaminen on kielenpuhujalle yhtä luonnollista, ellei jopa luonnollisempaa, kuin sen tuottaminen: jokainenhan on ensin kuulija ja vasta sitten puhuja. Toisin kuin puheen tuottamisessa helposti ulkoisen tarkastelun kohteeksi alistuvaa elimistöä ei ole. Tieteellisesti uskottavaan havainnon eli perseption tutkimukseen tarvitaan vähintäänkin muita tutkittavia kuin tutkija itse. Havaintokokein

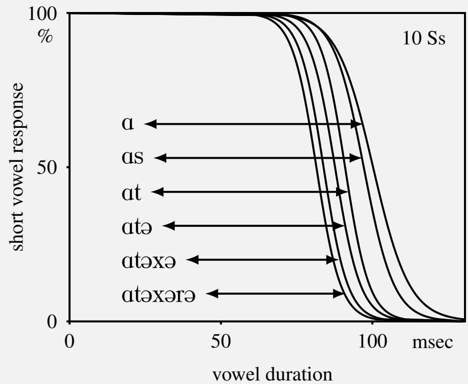
yritetään tavoittaa se, miten havainto syntyy ja mitkä tekijät siihen vaikuttavat, mitä voi olla mahdotonta tehdä pelkästään havainnoimalla tavallisia oikeita vastaanottotilanteita. Puheen havaitsemisen näkökulmaa on kutsuttu niin *perseptuaaliseksi* kuin *auditiiviseksi fonetiikaksi*.

Korvan tiedetään vastaanottavan äänisignaalin ja muuntavan sen toisenlaiseksi signaaliksi, mutta millään laitteitta tapahtuvalla tarkastelutavalla ei voi tarkkailla sitä, miten tämä tapahtuu. Käytännössä havaitsemista tutkitaankin usein havaintokokein eli yrittämällä selvittää muita kielenpuhujia tarkkaillen, mitä he havaitsevat ja mitkä seikat havaintoon voivat vaikuttaa.

**Tyypillisen havaintokokeen rakenne**. Tyypillisessä havaintokokeessa koehenkilöille soitetaan ärsykkeitä, minkä jälkeen he saavat kertoa, miltä kukin ärsyke kuulostaa. Yksinkertaisessa havaintokokeessa ärsykkeinä käytetään luonnollista, oikeiden puhujien ääntämää puhetta, esimerkiksi yksittäisiä sanoja, ja kuulijan tehtävänä on kertoa, minkä sanan hän kuulee. Näin voidaan selvittää, kuinka varmalla pohjalla jokin hypoteettinen oppositio tutkitussa kieliyhteisössä on. Tietenkin täytyy keksiä keino kertoa puhujille, mitä merkityksiä pyydetään tuottamaan, sekä vastaavasti keino kerätä kuulijoiden havaitsemat merkitykset. Jos kyseisellä kielimuodolla on yleisessä käytössä oleva kirjoitusjärjestelmä, joka johdonmukaisesti erottaa kyseiset merkitykset, tilanne on helppo. Mikäli oppositio on voimassa, siirrytään tyypillisesti jatkotutkimukseen, jossa aineistona toimivat käytetyt ärsykkeet, nyt puhujien ja kuulijoiden kategorisoinneilla täydennettynä, ja yritetään löytää akustisia piirteitä, jotka korreloivat merkityseron kanssa. Toinen yleinen havaintokokeen tyyppi on kahden kielen vertailu: toisen kielen puhujien tuotoksia voi soittaa toisen kielen kuulijoille, jolloin voi selvittää kielten välisiä eroja. Tällaisen kokeen tuloksena syntyy niin sanottu sekaannusmatriisi (engl. *confusion matrix*), jota voi analysoida erilaisilla tilastomenetelmillä.

Usein havaintokokeen ärsykkeet muodostavat jatkumoa edustavan sarjan, jossa vain yksi akustinen piirre muuttuu pienin täsmällisin askelin. Kyseisen akustisen piirteen arvellaan liittyvän olennaisesti tutkittavaan oppositioon. Koe paljastaa, onko näin, kuinka merkittävä akustinen piirre on opposition kannalta ja paljonko esiintyy variaatiota kuulijoiden kesken. Ärsykesarjan tekoon voi käyttää täysin syntetisoitua puhetta, jossa on helppoa kontrolloida akustiset erot tarkasti. Toisaalta, koska on

erittäin työlästä tuottaa havaintokokeisiin tarpeeksi laadukasta täysin synteettistä puhetta, toinen hyvin usein käytetty tapa on rakentaa ärsykkeet luonnollisesta (äänitetystä) puheesta siten, että yhtä ja samaa alkuperäistä äänitettä muokataan tarkasti muuttuvin askelin (ks. kuvio 5).




Esimerkki havaintokokeen tulosten esittelystä (identifikaatiofunktio; piirretty Nooteboomin kuvion mukaan, Nooteboom 1975, Figure 2.5.)

Kuvio 5. Esimerkki havaintokokeen tulosten esittelystä (identifikaatiofunktio, Nooteboom 1975).

Yksi ärsykesarja voi auttaa selvittämään, missä kulkee perseptuaalinen kategoriaraja ja kuinka jyrkkä se on, mutta tämä tieto sellaisenaan ei välttämättä ole kovin kiinnostava, ellei vertailla tuloksia toisiin vastaaviin kokeisiin. Esimerkiksi saattaa olla mahdollista rakentaa ärsykkeet siten, että sarjan ääripäät selvästi edustavat opposition kahta kategoriaa kahdessa vertailtavassa kielessä. Havaintokokeen toteuttaminen kummankin kielen kuulijoiden kanssa mahdollistaa kielten välisen vertailun.

Hyvin yleinen havaintokokeen muunnelma on myös sellainen, jossa on kaksi tai monta rinnakkaista ärsykesarjaa. Ärsykkeet ovat identtiset kaikissa sarjoissa yhtä eroa lukuun ottamatta. Kyseinen akustinen ero ei ehkä sellaisenaan riitä erottamaan merkityksiä luotettavasti, mutta se

saattaa esimerkiksi hiukan siirtää kategorioiden rajaa, mikä osoittaisi sen olevan relevantti havaitsemisessa. Esimerkiksi koesanan laajempi konteksti (kehyslauseen jokin ominaisuus, kuten tempo) saattaa toimia näin.

**Havainnon mittaaminen**. Helpoin ja yleisin tapa mitata havaitsemista on kysyä havaitsijalta, mitä hän on havainnut. Koehenkilön tehtävää, jossa tarkoitus on nimetä eli luokitella tarjottuja ärsykkeitä, kutsutaan *identifikaatiotehtäväksi*. Joskus tällaisessa tehtävässä koehenkilö saa vapaasti luokitella kuulemansa, mutta usein valintamahdollisuuksia rajataan etukäteen esim. kahteen tai kolmeen, ja koehenkilön on pakko valita yksi annetuista vaihtoehdoista jopa silloin, kun hänestä tuntuu, että valinta on täysin mielivaltainen. Tällainen pakkovalinta-asetelma saattaa olla teknisesti helpommin järjestettävissä ja sen tulokset saattavat olla helpommin tulkittavissa verrattuna tehtävään, jossa vastausvaihtoehtojen määrää ei ole rajattu.

Usein halutaan myös selvittää, millaisia puheessa esiintyviä eroja kuulijat yleensä pystyvät kuulemaan. Ero, jota ihminen ei pysty kuulemaan, ei luonnollisesti kelpaa kielellisen viestinnän käyttöön. Tehtävää, jolla yritetään selvittää ihmisen erottelukyvyn rajoja, kutsutaan *diskriminaatiotehtäväksi*. Yksinkertaisimmassa tapauksessa koehenkilö kuulee aina kaksi ärsykettä ja responssina kertoo, olivatko ne identtisiä vai erilaisia (kumpaakaan ei siis tarvitse luokitella), ja vaikka on kehitetty myös joukko hieman mutkikkaampia mutta luotettavampia koeasetelmia, tavoite on kaikissa sama. Erottelukykyyn vaikuttaa tietenkin ihmisen kuulomekanismi yleensä, mutta taustalla saattaa olla muitakin tekijöitä, kuten äidinkieli, ikä, tms. Myös erilaisilla aivosähkökäyrämittauksilla (herätevaste) on mahdollista nykyisin selvittää, mitä ääniä ihminen pystyy erottamaan erilaisissa olosuhteissa ilman, että häneltä tarvitsisi pyytää mitään aktiivista responssia (ks. Psyko- ja neurolingvistiikka tk.).

Samaa havainnon tarkkuutta ja luokittelua yritetään arvioida erilaisissa asetelmissa, joissa koehenkilön tehtävänä on toistaa tai matkia kuulemaansa. Ärsykkeinä voi toimia vaikkapa koehenkilölle tuntemattoman kielen ääntämyksiä tai muokattuja äidinkielisiä ääntämyksiä. Tällöin ehkä selviää, mitkä seikat jäävät vaille huomiota, mutta myös, miten oma kielikokemus vaikuttaa siihen, mitä hän kuulee.

Havaitsemisen vaikeutta eri tilanteissa voidaan arvioida esim. mittaamalla reaktioaikoja tehtävässä, jossa koehenkilö yrittää reagoida

ärsykkeisiin niin nopeasti kuin mahdollista. Myös silmänliikkeiden seuranta auttaa valaisemaan havaintoprosessin etenemistä tehtävässä, jossa havaittavat vaihtoehdot ovat koko ajan visuaalisesti läsnä (esim. eri sanoihin viittaavia kuvia).

**Havainnon manipulointimahdollisuuksia**. Pohjustustehtävällä (engl. *priming*) voidaan tutkia ennakko-odotusten vaikutusta puheen havaitsemiseen. Siinä koehenkilö saa havaita identtisiä ärsykkeitä eri kontekstiärsykkeiden jälkeen. Edeltävä kontekstiärsyke saattaa muuttaa koehenkilön havaintoa varsinaisesta tutkittavasta ärsykkeestä. Havaitsemisen ajallista ulottuvuutta voi tutkia esimerkiksi porrastetulla kuuntelulla (engl. *gating*) eli esittämällä puheen siten, että kuulija kuulee pienin askelin enemmän ja enemmän signaalia (ks. Grosjean 1980; 1996). Dikoottinen kuuntelu puolestaan on menetelmä, jossa koehenkilöllä kuuntelutetaan ääniärsykkeitä kuulokkeiden kautta, jotta vasempaan ja oikeaan korvaan voidaan soittaa eriävää signaalia. Jos halutaan tutkia esim. tarkkavaisuuden vaikutusta, vasen ja oikea kanava voivat olla täysin eri puhetta. Toisinaan taas sama puhesignaali on jaettu niin, että vasen ja oikea korva saavat eri osia samasta puheesta, jolloin ehkä selviää, tapahtuuko havaitseminen varhaisemmassa – ennen korvien signaalien yhdistämistä – vai myöhäisemmässä vaiheessa.

## 3.4. Muita aspekteja

**Puhujien rytmi ja synkronointi**. Pitkään on kiinnitetty huomiota siihen, että puheen avulla välitetään informaatiota. Viime aikoina on enenevässä määrin nostettu esiin puheen synkronoiva funktio. On selvä, että puhujan eri artikulaatioliikkeitä pitää tavalla tai toisella keskenään tahdistaa eli synkronoida eli kohdistaa toisiinsa ajallisesti. Myös viestinnän osallistujat tahdistuvat eli ikään kuin pääsevät samalle aaltopituudelle puheen avulla. Informaatio- ja synkronointifunktiot oikeastaan edellyttävät toisiaan: yhteinen puherytmi luo yhteisiä odotuksia, jotka sitten mahdollistavat yhteisiä valintoja, mutta valinnat myös ylläpitävät puhujien yhteistä rytmiä. Puhujien synkronointikykyä on tutkittu kekseliäin menetelmin, esimerkiksi simultaaniluennalla, jossa koehenkilöt lukevat samaa tekstiä

yhdessä (Cummins 2009). Myös äänitettyjä keskusteluaineistoja voi analysoida ääntämisen rytmin ja puhujien synkronoinnin kannalta. Pitkään vallinnut käsitys, että maailman kielet kuuluisivat rytmiltään vain kahteen luokkaan, ns. tavu- ja painoajoitteisiin kieliin, on osoittautunut virheelliseksi. Kielet kyllä eroavat rytmeiltään toisistaan, mutta puherytmi on kompleksisempi ilmiö ja sen kirjo laajempi kuin aiemmin on ymmärretty. Sen lisäksi kunkin kielen sisällä yksilöstä toiseen ja tilanteesta toiseen on paljon rytmivaihtelua, joskus jopa enemmän kuin kielten välillä. Toistaiseksi tätä aspektia on tutkittu verraten vähän.

**Sosiofonetiikka**. Sosiofonetiikka on ollut kasvava tutkimussuunta viime vuosikymmeninä (ks. esim. Celata & Calamai 2014). Ala voidaan määritellä joko sosiolingvistiikan ja fonetiikan sulautumaksi (Thomas 2002, 189) tai foneettiseksi tutkimukseksi, jossa tavallista enemmän otetaan huomioon viestinnän sosiaalinen luonne. Nimitys vakiintui käyttöön 1970-luvun lopulla. Sosiofoneettisia tutkimuksia yhdistäväksi punaiseksi langaksi voi nähdä sen, miten ja miksi puhe sosiaalisesti vaihtelee ja miten kielenpuhujat tätä vaihtelua viestinnässään hyödyntävät (Foulkes, Scobbie & Watt 2010, 704).

# 4. Yhteenveto

Edellä on esitelty esimerkkejä kokonaisista tutkimusasetelmista sekä yleisemmin tutkimusasetelman osia, mahdollisuuksia tehdä havaintoja puheesta ja toisaalta mahdollisia keinoja tarkasti kontrolloida puheen piirteitä. Esillä on ollut nimenomaan kielitieteellinen fonetiikka, mutta monien eri tieteiden – jopa laajempien tieteenalojen, kuten humanististen ja luonnontieteiden – rajalla sijaitsevana tieteenä fonetiikkaa voisi tarkastella monella muullakin tapaa. Se ulottuu esimerkiksi niin insinööritieteisiin (tutkimuskohteinaan esimerkiksi puhesynteesi tai automaattinen puheen tunnistaminen) kuin lääketieteeseenkin (jolloin kohteena voivat olla esimerkiksi puheen häiriöt). Yhdistävänä tekijänä on inhimillinen puhe, mutta tutkimusotteita siihen on lukemattomia.

**Aiheesta lisää:**


Aaltonen, Olli, Aulanko, Reijo, Iivonen, Antti, Klippi, Anu & Vainio, Martti (toim.). 2009. *Puhuva ihminen. Puhetieteiden perusteet*. Helsinki: Otava.

Hardcastle, William J., Laver, John & Gibbon, Fiona E. (toim.). 2010. *The Handbook of Phonetic Sciences*. Toinen laitos. Malden (MA): Wiley-Blackwell.

Iivonen, Antti (toim.). 2005. *Puheen salaisuudet. Fonetiikan uusia suuntia*. Helsinki: Gaudeamus.

Iivonen, Antti, Aulanko, Reijo & Vainio, Martti (toim.). 2005. *Monikäyttöinen fonetiikka*. Helsingin yliopiston fonetiikan laitoksen monisteita 21. Helsinki: Helsingin yliopisto.


## VIITTEET

1 ”[– –] ei puhuminen [– –] ole puhe-elinten määräasentojen jyrkkää vaihtelua, hyppäämistä toisen äänteen vaatimasta asemasta seuraavan vaatimaan, vaan [– –] pysähdyksittä jatkuvaa liikettä [– –]” (Penttilä 1948, 7).

2 Engl. *score*; esim. nk. artikulatorisen fonologian teoriassa.

3 Observoitavat muuttujat on toisinaan nimetty riippuviksi tai selitettäviksi muuttujiksi, ja kontrolloidut muuttujat on toisinaan nimetty riippumattomiksi tai selittäviksi muuttujiksi.

4 ”Meaning [– –] is to be regarded as a complex of contextual relations, and phonetics, grammar, lexicography and semantics each handles its own components of the complex in its appropriate context.” (Firth 1957 [1935], 19).

5 Laitteiston rakenteesta ja toiminnasta ks. https://www.youtube.com/watch?v=6oIe-joZ17jo.

6 Digitoija on analogi-digitaali- eli A-D-muunnin esimerkiksi tietokoneen äänikortissa. Tietokone käsittelee digitaalista dataa, mutta ääni akustisena ilmiönä on luonteeltaan analogista, joten se täytyy ensin muuntaa.

7 Myös *sonagrammi* vanhan tuotenimen mukaan.

8 Tarkalleen ottaen F0 ja havaittu äänenkorkeus ovat tietenkin eri asioita, joskin määräsuhteessa toisiinsa.

## KIRJALLISUUS


Ashby, Michael. 2017. The digital history of phonetic science. Julkaisussa: Vainio, Martti, Šimko, Juraj & Aulanko, Reijo (toim.) *Proceedings of the Second International Workshop on the History of Speech Communication Research: Helsinki, August 18–19, 2017*. Dresden: TUDpress, 7–15.

Benus, Stefan & Gafos, Adamantios I. 2007. Articulatory characteristics of Hungarian transparent vowels. *Journal of Phonetics* 35:3, 271–300.

Boersma, Paul & Weenink, David. 1996. *Praat, a system for doing phonetics by computer, version 3.4*. Technical report 132. Amsterdam: Institute of Phonetic Sciences of the University of Amsterdam.

Boersma, Paul & Weenink, David. 2018. Praat: doing phonetics by computer (version 6.0.39). [tietokoneohjelma]. [viitattu 2.3.2020]. Saatavissa: http://www.praat.org/.

Celata, Chiara & Calamai, Silvia (toim.). 2014. *Advances in Sociophonetics*. The Hague: John Benjamins.

Cummins, Fred. 2009. Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics* 37:1, 16–28.

Firth, J. R. 1957 [1935]. The technique of semantics. Julkaisussa: Firth, J. R. *Papers in Linguistics: 1934–1951*. Ilmestynyt alkuaan lehdessä *Transactions of the Philological Society*. Oxford: Oxford University Press, 7–33.

Foulkes, Paul, Scobbie, James M. & Watt, Dominic. 2010. Sociophonetics. Julkaisussa: Hardcastle, William J., Laver, John & Gibbon, Fiona E. (toim.) *The Handbook of Phonetic Sciences*. Toinen laitos. Malden (MA): Wiley-Blackwell, 703–754.

Fowler, Carol A. & Turvey, Michael T. 1980. Immediate compensation in bite-block speech. *Phonetica* 37, 306–326.

Grosjean, F. 1980. Spoken word recognition processes and gating paradigm. *Perception & Psychophysics* 28, 267–283.

Grosjean, F. 1996. Gating. *Language & Cognitive Processes* 11, 597–604.

Kelso, J. A. Scott, Vatikiotis-Bateson, Eric, Saltzman, Elliot L. & Kay, Bruce. 1985. A qualitative dynamic analysis of reiterant speech production: phase portraits, kinematics, and dynamic modeling. *The Journal of the Acoustical Society of America* 77:1, 266–280.

Linell, Per. 2005. *The Written Language Bias in Linguistics. Its Nature, Origins, and Transformations*. London & New York (NY): Routledge.

Nooteboom, Sieb G. 1975. Context effects in the perception of phonemic vowel length. 8th International Congress of Phonetics Sciences, Leeds, elokuussa 1975.

Penttilä, Aarni. 1948. *Suomen kielen äänne- ja oikeinkirjoitusoppi*. Porvoo: WSOY.

Thomas, Erik. 2002. Instrumental phonetics. Julkaisussa: Chambers, J. K., Trudgill, Peter & Schilling-Estes, Natalie (toim.) *The Handbook of Language Variation and Change*. Malden (MA) & Oxford: Blackwell, 168–200.

Äimä, Frans. 1918a. *Phonetik und Lautlehre des Inarilappischen: I: Beobachtungsphonetik und deskriptive Lautlehre*. Helsinki: Société Finno-ougrienne.

Äimä, Frans. 1918b. *Phonetik und Lautlehre des Inarilappischen: II: Instrumentale Messungen*. Helsinki: Société Finno-ougrienne.

# Luku 5 Lingvistinen etnografia toimintayhteisöissä

*Heini Lehtonen*
https://orcid.org/0000-0002-8726-0352
*Sari Pöyhönen*
https://orcid.org/0000-0003-1507-7206

**Mikä?**

Lingvistinen etnografia on metodologia, joka pyrkii ymmärtämään toimintayhteisöjen sosiaalista toimintaa niiden sisältä käsin ja osallistujien lähtökohdista. Näitä lähtökohtia ei oleteta, vaan osa etnografiaa on niiden selvittäminen. Etnografi siis tunnustaa oman ulkopuolisuutensa ja tutkimukseen osallistujien asiantuntijuuden heitä itseään koskevissa asioissa.

**Katso myös:**

Luku 3 Haastattelu aineistonkeruun metodina sosiolingvistiikassa

Luku 13 Kielten dokumentointi ja kieliopin kuvaus

Luku 16 Antropologinen lingvistiikka

Luku 21 Sosiolingvistiikka

Luku 22 Keskustelunanalyysi ja vuorovaikutuslingvistiikka

# 1. Johdanto

Kuvitellaan, että haluaisit tutkia kielen sosiaalisia ulottuvuuksia ja vuorovaikutusta yhteisöissä, joita et tunne entuudestaan. Kuitenkin niille ihmisille, joita haluaisit tutkia, vieras ympäristö olisi tuttuakin tutumpi: he ehkä asuvat siellä tai kohtaavat siellä joka päivä. Heidän kielensä ja se, miten he käyttävät sitä jäsentämään suhteitaan toisiinsa ja ympäristöönsä, on heille jokapäiväistä, mutta sinä et tiedä siitä juuri mitään, vaikka kieltä osaisitkin. Millaisia mahdollisuuksia sinulla olisi vastata tutkimuskysymyksiisi? Millaisia metodeja tarvitsisit?

Tässä artikkelissa esittelemme etnografisten metodien keinostoa kerätä ja hyödyntää sosiolingvististä aineistoa. Keskitymme ennen muuta sellaiseen etnografiseen tutkimukseen, jota viime aikoina on erityisesti Iso-Britanniassa kutsuttu *lingvistiseksi etnografiaksi* (ks. Rampton, Tustin, Maybin, Barwell, Creese & Lytra 2004) niissä sosiolingvistisissä lähestymistavoissa, jotka on nimetty esimerkiksi *vuorovaikutussosiolingvistiikaksi*, *globalisaation* tai *mobilisaation sosiolingvistiikaksi*, *moninaisuuden sosiolingvistiikaksi* (engl. *sociolinguistics of diversity*), *kriittiseksi sosiolingvistiikaksi* tai *soveltavaksi sosiolingvistiikaksi* (ks. Blommaert 2010; Madsen, Karrebæk & Møller 2015; Heller, Pietikäinen & Pujolar 2018). Yhteistä näille lähestymistavoille on tarkastella kielellistä variaatiota vuorovaikutuksessa sekä sitä, mitä merkityksiä vuorovaikutukseen osallistujat (fyysisissä tai virtuaalisissa tiloissa) antavat laajemmille yhteiskuntarakenteille, kuten koulutusjärjestelmälle tai maahanmuuttopolitiikalle.

Tämän teoksen artikkeleista lähimpänä lingvistisen etnografian metodeita ja teorioita ovat sosiolingvistiikkaa, antropologista lingvistiikkaa, keskustelunanalyysia ja vuorovaikutuksen tutkimusta sekä haastattelu- ja käsittelevät luvut. Lingvistisen etnografian keinoin tutkitaan usein juuri sosiolingvistisiä kysymyksiä, ja sillä on yhteisiä juuria variaationtutkimuksen, lingvistisen antropologian ja etnometodologisen keskustelunanalyysin kanssa. Näissä kaikissa kiinnitetään huomiota aineiston autenttisuuteen, eikä kokeellisia metodeja yleensä käytetä. Etnografisesti kerättyjä aineistoja tarkastellaan usein variaationtutkimuksen tai vuorovaikutustutkimuksen metodein, mutta etnografia ei kuitenkaan ole vain

aineistonkeruun metodi, vaan siihen sisältyy jatkuvasti myös analyysia ja tulkintaa (Blommaert & Jie 2010, 7–15).

Suomessa lingvistisen etnografian keinoin on viime aikoina tarkasteltu muun muassa kansainvälisissä työtehtävissä toimivien insinöörien kokemuksia englannin kielen käytöstä sekä heidän kielellistä repertuaariaan (Räisänen 2013), inkerinsuomalaisten vanhusten kielioloja (Mononen 2013), *autenttisuuden* käsitteen rakentumista suomalaisessa hip hop -musiikissa (Westinen 2014), nuorten kielellisiä resursseja ja niiden käyttöä sosiaalisessa asemoinnissa monietnisissä helsinkiläiskouluissa (Lehtonen 2015), osallistumisen ja identifioitumisen tiloja kristittyjen nuorten elämäntapaurheilijoiden yhteisössä (Peuronen 2017), toisen sukupolven aikuisten maahanmuuttajien kieli-identiteettiä (Bäckman 2017) sekä turvapaikanhakijoiden arkea ruotsinkielisellä Pohjanmaalla (Pöyhönen, Kokkonen & Tarnanen 2019). Virtuaalisen vuorovaikutuksen kasvaessa myös verkkoetnografiat ovat kehittyneet omaksi tutkimussuuntauksekseen (ks. Androutsopoulos 2014). Ne tutkivat kielellisten ja kulttuuristen innovaatioiden virtauksia verkkoympäristöissä. Esimerkkinä voidaan mainita Kytölän (2013) tutkimus suomalaisten jalkapalloaiheisten keskustelufoorumien vuorovaikutuksesta. Kaikille näille tutkimuksille yhteistä on, että ne tarkastelevat yhteisesti jaettujen ideologioiden ja sosiaalisten normien sekä yksilöiden käyttämien kielellisten resurssien leikkauspistettä toimintayhteisöjen käytänteissä.

Resurssi voi olla mikä tahansa kieleen tai muuhun sosiaaliseen merkkijärjestelmään liittyvä valinta, jonka avulla luodaan merkitystä vuorovaikutustilanteessa. Taustalla on ajatus kielestä osana sosiaalista semioosia. Visakon (2015, 22) mukaan semioosi on aikaan sidottu kudelma prosesseja, joissa merkkien tulkitaan merkitsevän jotakin ja näin tuotetaan esimerkiksi ”merkitystä”, ”tietoa” ja myös ”persoonia”.[1] Lingvistiset etnografiat taas kuvaavat niitä sosiaalisia mekanismeja, joilla semioottisia resursseja, mukaan lukien kielellisiä resursseja, omaksutaan, kierrätetään, nimetään ja tulkitaan uudelleen.

*Toimintayhteisön* käsite on hyödyllinen lähtökohta. Käsite tarjoaa jokapäiväisemmän ja helpommin operationalisoitavan vaihtoehdon abstraktille *puheyhteisön* käsitteelle (Lave & Wenger 1991; Eckert & McConnell-Ginet 1992; Wenger 1998; Meyerhoff 2002; Barton &

Tusting 2005). Siinä missä puheyhteisö on abstrakti ja kuviteltu (Anderson 1983), *toimintayhteisöllä* tarkoitetaan sellaisia yhteisöjä, joihin samat ihmiset kokoontuvat jonkin yhteisen asian ja yhteisen toiminnan äärelle. Toimintayhteisö voi olla esimerkiksi koulu, harrastuspiiri, työpaikka, kunnanhallitus, vastaanottokeskus tai periaatteessa mikä tahansa enemmän tai vähemmän yhteisten tavoitteiden ja sääntöjen mukaan toimiva ryhmä. Toimintayhteisöissä syntyy niiden omia käytänteitä ja normeja, toiminnan tapoja, ja kieli ja sen merkitykset elävät ja muuttuvat näissä käytänteissä ja toiminnoissa.

Vaikka tutkittava toimintayhteisö olisi suomalainen koulu ja etnografi olisi itse aiemmin käynyt suomalaisen koulun, hän ei voi olettaa läheskään kaikkia eri ajan ja paikan käytänteitä tutuiksi. Artikkelin alussa oletimme yhteisön, jossa etnografi ei ole koskaan ennen käynyt, mutta tämä ei tarkoita, että olisi lähdettävä tuhansien kilometrien päähän täysin toisenlaiseen kieli- ja kulttuuriympäristöön (ks. Kielten dokumentointi ja kieliopin kuvaus tässä kirjassa). Lingvistisen etnografian tutkimuskohteet voivat olla hyvinkin lähellä, ja yhä useammin ne ovat yhteisöjä, joiden toimintaan tutkija osallistuu. Tutkimuksen tavoitteena on nähdä tuttuuden ja itsestäänselvyyden läpi ja purkaa luutuneita tai suorastaan myytinomaisia arkisia käsityksiä kielestä ja sosiaalisista identiteeteistä (ks. Copland & Creese 2015), vaikkapa siitä, millaisille ihmisille suomen kieli kuuluu ja millaista suomen kieli voi olla. Silloin kun etnografi toimii yhteisössä, joka on hänelle omakohtaisesti tuttu tai jonka jäsen hän on, hän tasapainottelee jäsenyytensä ja osin ulkopuoliseksi tarkkailijaksi asettuvan tutkijan asemassa, mikä on tiedostettava analyysissa. Olipa kenttä alun perin tuttu tai vieras, etnografian tekeminen on joka tapauksessa aaltoliikettä, jossa välillä kokee osallisuutta ja järjestystä, välillä ulkopuolisuutta ja tietämättömyyttä.

Etnografian metodologiaan sisältyvät ainakin sen tietoteoreettiset taustakäsitykset, erilaiset kenttätyön ja aineistonkeruun metodit ja periaatteet, aineiston käsittelyn ja analyysin tavat, keinot välittää tutkimustuloksia erilaisille yleisöille ja takaisin tutkimusyhteisöön sekä tutkijanposition kriittinen tarkastelu. Etnografia ei siis ole vain yksi(n) metodi, eikä ole vain yhtä tapaa tehdä etnografiaa. (ks. myös Diskurssianalyysi

tk.) Kaikilla etnografioilla on kuitenkin joitakin yhteisiä piirteitä, joista luomme yleiskuvan.

Luvussa 2 esittelemme etnografian keskeisiä piirteitä ja sen kirjon kehittymistä sellaiseksi kuin se nykyisin on sekä erityisesti sen hyödyntämistä suomalaisessa sosiolingvistiikassa. Luku 3 käsittelee etnografin käytännön työtä aineistonkeruusta analyysiin ja tutkimuseettisiin valintoihin. Luku 4 on yhteenveto.

# 2. Tiheä kuvaus

## 2.1. Etnografian keskeiset piirteet

Vaikka etnografiassa ei voikaan määritellä yhtä ontologista, epistemologista tai metodologista taustaoletusta, se yhdistetään tavallisesti ymmärtävään tai tulkitsevaan lähestymistapaan (ks. Johdannon luku 2 tk.), jossa tutkimuksen tavoitteena on ymmärryksen lisääminen tutkittavasta ilmiöstä (Guba & Lincoln 1998; Atkinson & Hammersley 1998). Etnografian juuret ovat antropologiassa ja erityisesti tulkitsevassa, geertziläisessä kulttuuriantropologiassa (ks. esim. Ruohonen & Laitila 1994). Siinä pyritään kokonaisvaltaiseen tulkintaan, jossa kulttuurinsisäiset symbolit saavat merkityksen osana laajempaa kulttuurista kokonaisuutta, ei siitä irrallaan. Kulttuuri nähdään siis symbolisena systeeminä tai tekstinä, jonka merkityksiä luetaan, käännetään ja tulkitaan. Tulkitessaan tutkija kirjoittaa tiheän kuvauksen (engl. *thick description*, ks. Geertz 1973, 6), tekstin, jossa hän kuvaa yhteisön toimintaa ja kontekstia niin, että se vastaa yhteisön jäsenten kokemuksia ja on ulkopuoliselle ymmärrettävä. Jos tutkimukseen osallistujat eivät tunnista etnografin tekstiä, jokin on mennyt pahasti pieleen. Tutkija ei ole tällöin toiminut riittävän pitkään vuorovaikutuksessa tutkimukseen osallistujien kanssa, ei ole kuunnellut, mitä osallistujilla on ollut kerrottavanaan, eikä ole osannut liittää tutkimaansa toimintaa kontekstiin. Tekstistä on tullut tyhjä kuvaus.

Etnografista tutkimusta on Suomessa tehty usealla tieteenalalla, ja esimerkiksi kasvatus- ja sosiaalitieteissä kouluetnografialla on

pitkä traditio (Lappalainen, Hynninen, Kankkunen, Lahelma & Tolonen 2007; Gordon, Hynninen, Lahelma, Metso, Palmu & Tolonen 2007). Etnografian monitieteisyyden ja monimenetelmäisyyden vuoksi sitä on hankalaa määritellä tiiviisti (ks. myös Diskurssianalyysi tk.). Etnografia yhdistetään tavallisesti laadulliseen tutkimukseen, mutta tutkija voi hyödyntää myös määrällisiä menetelmiä, kuten laajoja rekisteriaineistoja. Vaikka etnografiaa voi tehdä monella tavalla, suurimpaan osaan etnografioista liittyy näitä piirteitä:

- (riittävän) pitkäaikainen läsnäolo kentällä
- pyrkimys kokonaisvaltaiseen analyysiin ja niin sanottuun tiheään kuvaukseen
- erilaisten aineistonmuotojen käyttö; aineiston triangulaatio
- tutkijanposition reflektointi
- tutkimustulosten monenlainen levittäminen
- tutkittavan yhteisön voimaannuttaminen tai osallistaminen.

## 2.2. Lingvistisen etnografian kehityskulkuja

Sosiolingvistisen etnografian juuria voi jäljittää toisaalta yhteiskuntatieteellisiin kaupunkietnografioihin (Deegan 2007), toisaalta lingvistisen antropologian kenttätyöperinteeseen ja puheen etnografiaan, jota alun perin käytettiin Pohjois-Amerikan alkuperäiskansojen kielten kuvaukseen (Hymes & Gumperz 1972; Duranti 1997; ks. myös Antropologinen lingvistiikka, Kielten dokumentointi ja kieliopin kuvaus tk.). Lingvistisen etnografian kannalta keskeinen on Dell Hymesin perintö (ks. Rampton 2007). Hymesin varhaiset, nyt klassikoiksi muodostuneet kirjoitukset (1972a; 1972b; 1974) olivat alun perinkin reaktiota Chomskyn generatiiviseen kielioppiin ja chomskylaiseen käsitykseen universaalista kielikyvystä sekä ideaalista yksikielisestä natiivipuhujasta. Hymes huomautti, että kaikissa kieliyhteisöissä esiintyy variaatiota ja monikielisyyttä ja kaikkien kielelliset kompetenssit ovat toisistaan poikkeavia – ideaalia yksikielistä natiivipuhujaa on siis vaikea löytää (1972a). Yksikielisyyden normin sekä *yhtenäisen kielimuodon* ja *natiivipuhujan*

käsitteiden kyseenalaistaminen ovat yhä lingvistisen etnografian keskeisiä teemoja.

Hymesin ja tämän kollegan, vuorovaikutussosiolingvistiikan pioneerin John Gumperzin kehittämä puheen etnografia (myöhemmin myös viestinnän etnografia) pyrki saamaan kiinni nimenomaan kielen kontekstuaalisuudesta; siitä, että kielellinen merkitys on aina kiinni ajassa, paikassa ja tilanteisessa kontekstissa. Hymesin ja Gumperzin perintö näkyy uudemmassa lingvistisessä etnografiassa muun muassa siinä, että vuorovaikutusta kuvataan usein sellaisissa yhteisöissä, joissa osallistujien kielelliset resurssit ovat jollain lailla epäsymmetrisiä ja vuorovaikutukseen liittyy jatkuvaa neuvottelua valtasuhteista. Monet ovat kuitenkin huomauttaneet uudemman lingvistisen etnografian eroavan lingvistisen antropologian perinteestä siten, että siinä missä lingvistisessä antropologiassa on lähdetty kauas ja vieraiden pariin, lingvistisen etnografian tutkimuskohteet ovat usein lähellä ja näennäisen tuttuja (Hammersley 2007, 689; Copland & Creese 2015, 13–14).

Hymesilainen traditio ja variaationtutkimus kohtaavat niin sanotussa kolmannen aallon variaationtutkimuksessa (Eckert 2012; variaationtutkimuksen kolmesta aallosta ks. myös Sosiolingvistiikka tk.), jossa ollaan kiinnostuneita kielenpiirteiden sosiaalisesta indeksisyydestä eli siitä, millaisten ideologisten ja sosiaalisten prosessien kautta merkkejä liitetään tietynlaisiin ihmisiin ja miten merkeillä jäsennetään ihmistenvälisiä suhteita. Siinä missä ensimmäisessä ja toisessa aallossa variaatiota tarkastellaan *variaabelin* ja *variantin*[2] käsitteiden avulla, kolmannen aallon tavoitteena on kuvata minkä tahansa kielenpiirteen sosiaalista merkitystä vuorovaikutuksessa.

*Resurssi*-termi on yleistynyt vaihtoehtona *variantin* käsitteelle. Quistin (2005, 204–205) mukaan kielellinen resurssi on paikallista ja sosiaalista merkitystä kantava ja luova kielellinen mahdollisuus, jota puhujat voivat aktiivisesti käyttää tietynlaisen persoonan luomiseen ja itsen asemointiin. Jørgensen ja Møller (2014) ovat käyttäneet myös termiä *piirre* (engl. *feature*) korostaakseen sitä, että kielen sosiaalisia ulottuvuuksia aletaan selvittää etnografisesti merkittäväksi osoittautuvasta kielenilmiöstä lähtien eikä ennalta määritellystä variaabelista tai varieteetista käsin (ks. myös Lehtonen 2015, 37–38; Westinen 2014, 129–130).

Yksinkertaistettuna tämä tarkoittaa sitä, että kun etnografi-lingvisti menee kentälle, hänellä ei ole mielessään tiettyä kielenpiirrettä (tai variaabelia), jota hänen on tarkoitus tutkia, vaan hän saa vasta havainnoimalla selville, mitkä resurssit vuorovaikutuksessa osoittautuvat merkityksellisiksi tutkittavan yhteisön sosiaaliselle järjestykselle.

Variaationtutkimuksen kolmas aalto siis käsitteellistää variaatiota toisin kuin labovilaisen *variaabelin* käsitteen kautta. Samoin kyseenalaistetaan kielimuoto tai lekti[3] tarkastelun lähtökohtana (ks. Quist 2008). Kieli nähdään ennemmin toimintana kuin oliona: tarkastelun kohteena on se, millaisia merkityksiä milläkin piirteellä tietyssä tilanteessa ja toiminnassa herätetään (ks. Johdannon luku 3 tk.). Piirteet liittyvät rekistereihin, jotka voivat olla ”kieliä” (kuten suomi, englanti, havaiji) mutta myös muunlaisia sosiaalisen tulkinnan ja historiallisten prosessien kautta hahmonsa saavia kokonaisuuksia (kuten hip hop -tyyli, opetuspuhe) (Agha 2007; Madsen, Karrebæk & Møller 2015). Tutkimuksen kohteita ei siis lähtökohtaisesti oleteta jonkin aiemmin kuvatun kielimuodon osaksi, vaan tarkastellaan juuri niitä sosiaalisia prosesseja, joissa tulkinnat kielenpiirteistä ja kielimuodoista tutkitussa yhteisössä syntyvät.

Kolmas aalto tarkastelee siten yksilön, ryhmän, toimintayhteisön ja laajemman yhteiskunnallisen kontekstin suhdetta toisiinsa sekä niitä mekanismeja, joilla kielenpiirteen sosiaalinen merkitys syntyy, muokkautuu, uusiutuu ja vahvistuu vuorovaikutuksessa. Tarkastelun metodologiana on lingvistinen etnografia toimintayhteisössä. Vaikka kaikilla laadullista sosiolingvistiikkaa tekevillä tutkijoilla ei ole taustaa variaationanalyysissa, kolmannen aallon myötä niin labovilaisen koulutuksen saaneet tutkijat kuin myös muita teitä vuorovaikutussosiolingvistiikkaan saapuneet alkoivat esittää kysymyksiä, joihin vastaamisessa etnografinen lähestymistapa on keskeinen. Nuo kysymykset liittyvät juuri yksilön ja paikallisen yhteisön kielen variaation, sosiaalisen asemoinnin sekä laajemmin jaettujen kieli-ideologioiden yhteyksiin. Kolmen aallon metafora on harhaanjohtava siinä, että se saattaa synnyttää kronologisuuden vaikutelman tai sellaisen käsityksen, että uusi aalto pyyhkisi vanhan pois. Näin ei kuitenkaan ole: aallot rakentuvat limittäin ja toistensa varaan.

Tutkijan oman kielikäsityksen hahmottaminen on välttämätöntä tutkimuksen kannalta. Ei voi kuvata mitään, jollei selvitä itselleen

kuvattavan kohteen perusolemusta ja niitä olettamuksia, joista kuvauksen aloittaa. Lingvistisessä etnografiassa kieleen liittyviä ontologisia kysymyksiä sekä tutkijan positiota onkin pohdittu verrattain paljon (ks. Heller 2011; Madsen, Karrebæk & Møller 2015). Näitä voisi pitää myös sen tutkimuskohteina. Lingvistisen etnografian ja uudempien laadullisten sosiolingvististen lähestymistapojen kielikäsitystä voi useimmiten luonnehtia holistiseksi, kontekstuaaliseksi, sosiokulttuuriseksi ja dialogiseksi. Viime aikoina esillä ovat olleet myös lingvistisen antropologian piirissä syntyneet sosiaalisen indeksisyyden ja rekisteriytymisen teoriat (Ochs 1992; Silverstein 2003; Agha 2007; Woolard 2008). Erittelemme seuraavaksi lyhyesti näitä lähtökohtia ja käsitteitä.

**Holistisuus ja kontekstuaalisuus**. Etnografi kuvaa kielen limittäisyyttä muun toiminnan kanssa. Etnografiassa pyritään kuvaukseen tutkittavan yhteisön merkeistä ja merkityksistä, joiden avulla rakennetaan ja ylläpidetään yhteisön sosiaalista järjestystä. Kielelliset merkit eivät toimi tässä yksin, vaan merkitystä on ainakin toiminnalla, eleillä, ruumiillisuuteen ja ulkonäköön liittyvillä seikoilla ja kaikenlaisilla tyylillisillä ja elämäntapavalinnoilla. Tätä tarkoitetaan *holistisuudella* (vrt. Auer 2007, 11–15; Coupland 2007, 29–30; Quist 2005; Atkinson, Delamont & Housley 2008; Eckert 2012; Rampton, Maybin & Roberts 2015). Kaikenlaiset sosiaalisessa asemoinnissa hyödynnettäviä erontekoja tuottavat merkit ovat osa sosiaalista semioosia (ks. Visakko 2015, 22–23). *Puheen kontekstualisointi* taas on vuorovaikutussosiolingvistiikan keskeinen käsite. Sillä tarkoitetaan sitä, miten kielelliset valinnat herättävät assosiaatioita johonkin sosiaalisen maailman ilmiöön ja näin luovat ja muokkaavat sitä kontekstia, jossa juuri sanottua tulkitaan (ks. esim. Gumperz 1982; 1992; Auer 1992).

**Sosiokulttuurisuus ja dialogisuus**. *Kielen sosiokulttuurisuus* ja *dialogisuus* tarkoittavat sitä, että kielen ymmärretään syntyvän, elävän ja muuttuvan sosiaalisessa vuorovaikutuksessa. Moni tutkija on nostanut periaatteekseen Silversteinin (1985, 220) ”kaikenkattavan lingvistisen tosiseikan” (engl. *the total linguistic fact*)[4]. Sen mukaan kieli saa jatkuvasti uusiutuvat merkityksensä tilanteisessa kontekstissa mutta ihmismielessä ja kuitenkin osana kulttuurisia ideologioita (Lehtonen 2015, 54; ks. myös Woolard 2008; Blommaert & Rampton 2011; Madsen, Karrebæk

& Møller 2015). Puhujat kierrättävät tarjolla olevia kielellisiä resursseja ja tulkitsevat ja kontekstualisoivat niitä aina uudelleen uusissa tilanteissa. Lingvistisessä etnografiassa tai niin sanotussa moninaisuuden sosiolingvistiikassa ei siis pyritä yksittäisten kielten strukturaaliseen kuvaukseen. Kuvauksen kohteena on se, miten ”kieli tapahtuu” puhujien välillä, ja kielellisten merkkien merkitykset ovat ikään kuin olemassa vain osana sitä toimintaa ja vuorovaikutuskontekstia, joissa niitä kulloinkin käytetään.

Näissä lähestymistavoissa on problematisoitu myös eri kielten rajojen olemusta, kun aineistot selvästi osoittavat, että vuorovaikutuksessa vaivattomasti ylitetään eri kielten rakenteelliset rajat ja sosiaalisiakin tulkitaan uudelleen: vuorovaikutuksessa käydään eksplisiittistä ja implisiittistä neuvottelua siitä, mihin ja kenelle tietyt resurssit ”kuuluvat”. Puhujat käyttävät samassakin keskustelussa ”paloja” eri kielistä (engl. *bits of language*, ks. Blommaert 2010), ja samassakin keskustelussa käytetyt resurssit assosioituvat erilaisiin rekistereihin, esimerkiksi eri kieliin tai tyyleihin. Juuri lingvistisen etnografian piirissä onkin alettu etsiä vaihtoehtoisia tapoja käsitteellistää kielen moninaisuuteen ja limittäisyyteen liittyviä ilmiöitä. Tällainen on esimerkiksi ajatus ihmisestä kieleilijänä ja kielellisestä vuorovaikutuksesta kieleilynä (engl. *languaging*, ks. Jørgensen, Karrebæk, Madsen & Møller 2011; Madsen, Karrebæk & Møller 2015). Ajatus lähtee siitä, että ihminen ensisijaisesti kommunikoi niillä resursseilla, joita hänellä on käytössään ja jotka kussakin tilanteessa ovat toimivia ja sopivia ja osallistujien normiodotuksiin mukautuvia. Kielellinen ja kulttuurinen identifikaatio ja resurssien järjestäytyminen eri kieliin tai muihin rekistereihin kuuluviksi on sekundääriä ja ideologisen ja sosiaalisen tulkintaprosessin tulos. Rekisterit ja rekisteriytyminen ovat kuitenkin merkittävä osa tutkimuskohdetta sinänsä. Kielten sosiaalisilla rajoilla ja kielten nimeämisellä on merkitystä sosiaalisen asemoinnin kannalta, ja kielten rajoista käydyn neuvottelun tutkiminen on lingvistisen etnografian ydinaluetta.

**Sosiaalinen indeksisyys**. Lingvistinen etnografi kerää aineistoa, jonka avulla voidaan tarkastella semioottisten merkkien, vuorovaikutuksen, yksilön, toimintayhteisön sekä laajempien ideologioiden suhdetta toisiinsa. Linkkien luomisessa auttavat *sosiaalisen indeksisyyden* käsite ja rekisteriytymisen teoria (Agha 2007; suomeksi ks. Lehtonen 2015).

*Sosiaalisella indeksisyydellä* tarkoitetaan sitä, että merkillä, esimerkiksi kielenpiirteellä, on mahdollisuus herättää tietynlaisiin ihmisiin, heidän ominaisuuksiinsa tai ihmisten välisiin suhteisiin liittyviä assosiaatioita. Näin merkki toimii kontekstualisoinnin keinona, joka ohjaa myös merkin, puheen, toiminnan ja koko vuorovaikutustilanteen tulkintaa: kielellisen merkin kontekstuaalinen tulkinta sisältää käsityksen siitä, kuka puhuu, kenen äänellä ja missä asemassa. Lehtonen (2015; 2016) on tutkinut, miten tietyillä kielenpiirteillä luodaan puheeseen mielikuva ei-natiivista suomenpuhujasta. Tällaisilla *mina ei suami puhu* -lausahduksilla tuodaan vuorovaikutukseen esimerkiksi jonkinlaisen stereotyyppisen maahanmuuttajan ääni, mutta Lehtonen osoittaa, että niillä voidaan indeksoida myös ominaisuuksia, jotka ideologisesti ja stereotyyppisesti yhdistetään maahanmuuttajiin tai ulkomaalaisiin, esimerkiksi kyvyttömyyttä osallistua vuorovaikutustilanteeseen tai reagoida siihen. Tätä sosiaalista indeksisyyttä puhujat voivat hyödyntää asemoidessaan itseään meneillään olevaan vuorovaikutustilanteeseen.

Sosiaalinen indeksisyys syntyy rekisteriytymisen prosessissa. Rekisteriytyminen on aikaan ja paikkaan sidottu refleksiivinen prosessi (Agha 2007, 15–83). Siinä merkki saa tulkintoja, jotka mahdollistavat sen käytön sosiaalisessa asemoinnissa ja puheen kontekstualisoinnissa. Uusia mahdollisia merkityksiä syntyy vanhojen varaan, lomaan ja rinnalle. Rekisteriytyminen tapahtuu sosio-historiallisessa merkin käyttökertojen ketjussa, jossa merkkejä käytetään ja tulkitaan aina uudelleen. Esimerkki rekisteriytymisprosessin kuvauksesta on Vaattovaaran ja Halosen (2015) kuvaama niin sanotun stadilaisen ässän myytin synty: sen sosiaalinen merkitys on syntynyt vuosikymmenten aikana mielikuvista, jotka liittyvät tietyn helsinkiläisen tyttökoulun oppilaisiin, ensimmäisiin suomenkielisiin teatterinäytöksiin, iskelmätähtiin ja myöhemmin itähelsinkiläiseen tyttöduoon, viimeisimpänä homoseksuaalien urbaaniin elämäntapaan. Vaikka kaikki nämä merkitykset eivät jokaisessa merkin käyttötilanteessa välttämättä aktivoidu, ne ovat potentiaalisesti läsnä merkin kulloisessakin tulkinnassa. Eckertin (2008) käsittein ne ovat osa merkin indeksistä kenttää.

Lingvistisen etnografian ja kolmannen aallon variaationtutkimuksen metodologian kannalta keskeistä on, että rekisteriytymistä ei voi kuvata

vain kieliaineksen vaihtelua kuvaamalla, vaan on kuvattava myös kieliaineksen refleksiivinen tulkintaprosessi, siis yksinkertaisesti sanottuna se, millaisia merkityksiä kielenpiirteisiin liitetään. Toisin sanoen rekisteriytymiseen liittyy aina ideologinen tulkinta, joka syntyy vuorovaikutuksessa. Tähän tarvitaan niin sanottua metapragmaattista aineistoa, siis kieltä kommentoivaa kieltä tai toimintaa. Käytännössä voi tarkastella esimerkiksi jaksoja, joissa ihmiset puhuvat kielestä, tai sitä, millaista toimintaa herättää sellainen vuorovaikutusjakso, jossa kieli tavalla tai toisella nousee osallistujien huomion kohteeksi (ks. myös Lehtonen 2015, 41–42). Se, miten ihmiset mieltävät kielen ja sosiaalisten suhteiden yhteydet, ja se, miten ideologiat ohjaavat toimintaa ja siten vaikuttavat kielelliseen vaihteluun, ovat yhtä suuri osa kielen vaihtelun ja muutoksen mekanismia kuin vaihtelu sinänsä: jos kuvaa vain sitä, millaista vaihtelua aineistossa on, on oikeastaan kuvannut vasta osan siitä toiminnasta, joka vaihtelua aiheuttaa.

# 3. Kentälle ja kirjaan!

## 3.1. Aineistonkeruun käytänteitä

Etnografiselle tutkimusprosessille on tyypillistä teorian, tutkijan, tutkimukseen osallistujien, aineistonkeruun ja analyysin vuorovaikutteisuus. Tämä ilmenee esimerkiksi siten, että tutkija tavallisesti rajaa teoreettisen viitekehyksen ja asettaa itselleen lopullisen tutkimustehtävän vasta perehdyttyään tutkimukseen osallistujien kontekstiin ja heidän näkökulmastaan merkityksellisiin kysymyksiin aineiston keruun ja analyysin edetessä. Olennaiset kysymykset muotoutuvat usein vasta kenttätyössä. Esimerkiksi Jussi Lehtonen ja Sari Pöyhönen (2019) lähtivät liikkeelle Toinen koti -projektin alussa siitä, miten pakolaistaustaiset taiteilijat ja Suomessa syntyneet taiteilijat muodostavat ilmaisuyhteisön. Projektin edetessä merkitykselliseksi tuli kysyä, miten taiteen poliittisuuteen liitetyistä merkityksistä ilmaisuyhteisössä neuvotellaan teatteriesitystä rakennettaessa.

Etnografista tutkimusta tehdään yhä useammin monipaikkaisesti ja tutkimusryhmässä. Tyypillisesti *monipaikkaisella etnografialla* (Marcus 1995) tarkoitetaan saman ilmiön tutkimusta erilaisissa konteksteissa. Varsin usein tällaisessa tutkimuksessa on vertaileva näkökulma. Yhdenkin etnografian sisällä voi olla useita fyysisiä ja virtuaalisia tiloja, joissa tehdään tutkimusta. Esimerkiksi vastaanottokeskuksen etnografiassa (Pöyhönen, Kokkonen & Tarnanen 2019) tehtiin tutkimusta kahdeksassa paikassa: aikuisten turvapaikanhakijoiden asunnoissa, vastaanottokeskuksen toimistossa ja suomen kielen opetustiloissa, Oravaisten kylänraitilla, suomenkielisessä koulussa, alaikäisten ryhmäkodissa, ruotsinkielisessä kansanopistossa ja sosiaalisessa mediassa. Nämä kaikki tilat yhdistyivät samaan tutkimukseen tutkittavan ilmiön ja toimintayhteisön jäsenten kautta. Tutkimuksen ydinryhmään kuului kolme tutkijaa, joista yksi keskittyi aikuisten turvapaikanhakijoiden arkeen, yksi alaikäisten lasten koulunkäyntiin ja yksi työntekijöiden ammatillisuuteen. Riittävän pitkäkestoisen etnografian, yhteisen tutkimusaineiston ja kenttähavaintojen metakommentoinnin ja vinjettien (ks. Creese, Takhi & Blackledge 2016) eli tutkijoiden henkilökohtaisia kokemuksia purkavien kenttätekstien avulla tutkijaryhmä pystyi paremmin tavoittamaan kokonaiskuvaa tutkittavasta ilmiöstä. Tutkijoiden työnjako ja yhteistyö oli tärkeää myös vertaistuen ja työssä jaksamisen näkökulmasta, sillä tutkimusaihe ja avainhenkilöiden elämäntilanteet olivat varsin haastavia.

Mistä etnografinen aineistonkeruu sitten alkaa? Tyypillisesti etnografisen aineistonkeruun oppaissa keskitytään kentällä oloon eli tutkijan toimintaan yhteisössä. Tämä on ymmärrettävää, sillä aineistonkeruu on keskeinen osa mitä tahansa empiiristä tutkimusta. Etnografisen tutkimusprosessin voi myös hahmottaa kolmeen vaiheeseen: ennen kenttää, kentällä ja kentän jälkeen. Vaiheet eivät ole selvärajaisia, mutta niissä on tiettyjä painotuksia, joista tutkijan on hyvä olla tietoinen.

Etnografinen aineistonkeruu alkaa jo siinä vaiheessa, kun tutkija tutustuu toimintayhteisöön esimerkiksi kirjallisten dokumenttien kautta. Samalla hän opettelee yhteisön (esim. koulu, työpaikka, sairaala, vastaanottokeskus) kielellisiä ja kulttuurisia toimintatapoja, jotka ovat tärkeitä tutkimusprosessin kannalta. Etnografiassa harvoin asetetaan

hypoteesia, jota pyrittäisiin todistamaan oikeaksi tai vääräksi. Mielessä voi ennen kentälle pääsyä olla aihe tai fokus, johon etnografi haluaa keskittyä, mutta useimmiten on tärkeää aluksi saada käsitys siitä, mikä yhteisön vuorovaikutuksessa on ylipäätään merkityksellistä.

Ennen kentälle pääsyä tutkijan on myös selvitettävä toimintayhteisön portinvartijat: kenen kanssa tutkimuksesta on keskusteltava, jotta tutkimus olisi ylipäätään mahdollista? Tässä vaiheessa on myös erityisen tärkeää, että tutkija ei pitäydy omassa tutkijankielessään vaan kykenee kertomaan portinvartijoille tutkimuksesta niin, että toimintayhteisössä herää kiinnostus ja luottamus tutkimusta ja tutkijaa kohtaan. On hyvä oppia ilmaisemaan, jos ei ymmärrä jotain ilmiötä tai käytännettä, sillä tämä avaa toimintayhteisön jäsenille mahdollisuuden tuoda esille oma asiantuntemuksensa – heidän elämästäänhän on kysymys.

Portinvartijoiden kautta tutkijalla on mahdollisuus tutustua muihin yhteisön jäseniin. Tämä on tärkeä vaihe kaikessa etnografisessa tutkimuksessa mutta erityisesti silloin, kun tutkimuksessa on tarkoitus osallistaa toimintayhteisön jäseniä analyysiin ja tutkimuksesta viestintään (Bradley, Moore, Simpson & Atkinson 2018). Vuorovaikutussuhde tutkijan ja tutkimuksen osallistujan välillä taas edellyttää halukkuutta vuorovaikutukseen molemmilta osapuolilta. Tilanteet, joissa tutkimukseen osallistujat valikoituvat, ovat osin tietoisia tutkijan valintoja, osin tutkimuksen edetessä tapahtuvia sattumia. Etnografiseen prosessiin mahtuu yllätyksellisyyttä.

Triangulaatio on ikään kuin sisäänrakennettu etnografiaan (ks. Johdannon luku 4; Kielikontaktien tutkimus tk.). *Triangulaatiolla* tarkoitetaan sitä, että samaan havaintoon ja tulokseen voidaan päätyä erilaisten aineistojen kautta. Tällöin havainnot tukevat toisiaan ja lisäävät aineiston todistusvoimaa. Etnografialle ominaisia aineistonkeruun tapoja ovat (osallistuva) vuorovaikutustilanteiden havainnointi, keskustelut ja haastattelut osallistujien kanssa, spontaanin vuorovaikutuksen ääni- ja videotallenteet sekä moninaiset dokumentit ja artefaktit, kuten valokuvat, piirrokset, koulutehtävät, asiakirjat ja sosiaalisen median tekstit. Ilman riittävän pitkäaikaista ja systemaattista havainnointia ei ole etnografiaa, ja pelkkä havainnointi ei ole etnografiaa. Esittelemme seuraavaksi lyhyesti erilaisten aineistojen keräämiseen liittyviä kysymyksiä.

Tutkija kirjoittaa havainnoistaan muistiinpanoja ja kenttäpäiväkirjaa (ks. esim. Emerson, Fretz & Shaw 2007 [2001]). Kenttäpäiväkirjan ei tarvitse olla pelkkää tekstiä, vaan siihen voi yhdistää ääntä ja kuvaa. Näin muistiinpanot on helpompi yhdistää muihin aineistoihin, kun niihin on jo valmiiksi ankkuroitu tiettyjä hetkiä äänellisesti ja kuvallisesti. Etenkin pitkäaikaisessa etnografiassa monimediainen kenttäpäiväkirja on hyödyllinen. Tutkija voi kirjoittaa myös erillistä tutkimuspäiväkirjaa, jossa hän yhdistää analyysinsa kentällä tehdyistä havainnoista muihin tutkimusaineistoihin ja tutkimuskirjallisuuteen. Jotkut etnografit menevät kentälle havainnointia tukevan kaavion, taulukon tai muistilistan kanssa, toiset taas tekevät muistiinpanoja spontaanimmin sellaisista episodeista, joihin kiinnittävät huomionsa. Aina ei muistiinpanoja tehdessään osaa ennustaa, mikä on myöhemmän analyysin kannalta merkityksellistä. Siksi muistiinpanojen tekeminen on olennaista, ja siihen on hyvä kehittää itselle sopivia rutiineja.

Havainnointi ja kenttäpäiväkirja ovat toimivia tapoja aloittaa aineistonkeruu. Haastatteluiden avulla voi tutustua ihmisiin, ja ne koetaan myös luontevaksi tutkijan toiminnaksi. Etnografinen haastattelu on omanlaisensa tutkimushaastattelun laji: siinä tutkijalla ja tutkittavilla on yleensä enemmän yhteistä maaperää kuin haastattelututkimuksessa (ks. Sherman Heyl (2007 [2001]; ks. myös Haastattelu aineistonkeruun metodina sosiolingvistiikassa tk.). Haastattelu limittyy muuhun kenttätyöhön. Toisaalla alkaneet keskustelut voivat jatkua haastattelussa, tai haastatteluihin voidaan palata myöhemmin kenttätyön aikana. Haastatellessaan tutkija kartuttaa ymmärrystä siitä, mikä toimintayhteisön jäsenille on merkityksellistä. Haastattelut toimivat siten sekä tutkimuksen osallistujalle että tutkijalle analyyttisena taukona, jossa pohditaan toimintayhteisön käytänteitä. Haastattelut ovat myös hyvä silta vuorovaikutustilanteiden ääni- ja videotallenteisiin, jotka ovat olennainen osa lingvististä etnografiaa. Etnografisessa tutkimuksessa voidaan käyttää myös määrällisiä aineistonkeruu ja -analyysimenetelmiä tutkimuksen kontekstualisoimisessa tai yhdistämisessä vuorovaikutusanalyysiin.

Spontaanien vuorovaikutustilanteiden tallenteita on monenlaisia, ja niitä suunnitellessa on otettava huomioon ainakin seuraavia asioita:

- Miten tallennustilanne esitellään tutkimuksen osallistujille? Miten he ilmaisevat suostumuksensa (esim. suullisesti, kirjallisesti)?
- Millaista tekniikkaa tarvitaan: riittääkö tilanteessa yksi tallennuslaite vai tarvitaanko useampia?
- Liikkuvatko ihmiset tilanteessa: ovatko tallennuslaitteet paikoillaan vai liikkuvatko ne ihmisten mukana?
- Millä tavalla ja missä määrin tutkija on itse mukana tallennetuissa tilanteissa?
- Kuinka tärkeää on tallentaa koko tilassa tapahtuva toiminta videolle?
- Käytännön järjestelyissä on aina ajateltava sitä, että tallennuksesta on saatava selvää ja se on pystyttävä litteroimaan edes karkeasti. Jos puhujia on paljon eivätkä he ole tuttuja litteroijalle, video on yleensä tarpeellinen.
- Tutkimusympäristön institutionaalisuus vaikuttaa siihen, millaisia lupia tarvitaan.
- Tutkimusluvat saattavat aiheuttaa päänvaivaa, jos toimitaan paikassa ja tilassa, jossa liikkuu myös muita kuin tutkimukseen osallistuvia ihmisiä: Miten heidät rajataan pois tallenteista? Miten heille kerrotaan tutkimuksesta?
- Millainen tallennusformaatti on tutkimuksen kannalta käytännöllisin? Mitkä tiedostomuodot ovat yhteensopivia analyysivaiheessa käytettyjen ohjelmien kanssa? Miten ne tukevat monikielistä ja multimodaalista analyysia (eri kirjoitusjärjestelmät, kirjoitussuunnat, liike, kuva)?
- Miten aineisto säilytetään tietoturvallisesti? Miten se anonymisoidaan ja millaiset luvat sen jatkokäyttöön on? Mikä aineisto hävitetään ja mikä arkistoidaan?

Erilaisia tallennusvaihtoehtoja voi havainnollistaa esimerkiksi Lehtosen (2015) tutkimuksen avulla. Hän käytti luokkatilanteiden tallentamiseen videokameraa ja siihen ulkoista mikrofonia, joka asetettiin niiden

tutkittavien läheisyyteen, joiden vuorovaikutuksesta hän sillä hetkellä oli erityisen kiinnostunut. Lisäksi luokkatilassa saattoi samaan aikaan olla pari muuta nauhuria. Lehtonen antoi lisäksi tutkittavien kantaa mukanaan pientä äänitallenninta. Näin syntyi keskeinen osa tutkimuksen aineistosta. Tilanne, jossa tutkittavaa pyydetään päästämään tutkija – tallennuslaitteen kautta – lähelle, on hyvin herkkä. Tutkittavien on siinä vaiheessa jo jossain määrin luotettava tutkijaan. Vaikka tutkija ei olisi kaikissa tallennustilanteissa fyysisesti paikalla, hän on läsnä: hänelle saatetaan esimerkiksi puhua tallennuslaitteen välityksellä. Tämä ei missään nimessä tee aineistosta epäonnistunutta. Tutkijanposition jatkuva reflektointi on olennainen osa etnografiaa, ja kukin vuorovaikutustilanne on omalla tavallaan edustava. Juuri tallennustilanne saattaa suorastaan lisätä metapragmaattisen aineiston kannalta kiinnostavan ulottuvuuden, sillä se kannustaa tietynlaisiin kieltä kommentoiviin performansseihin. Tämä ilmenee myös esimerkiksi silloin, kun tutkimukseen osallistujaa pyydetään ottamaan valokuvia vuorovaikutuskäytänteistään ja ympäristöistään. Tutkimukseen osallistuja kokee tarpeelliseksi sanoittaa ottamiaan valokuvia ja siten perustella tilanteiden ja ympäristöjen merkitystä elämässään.

Etnografin aineisto on käytännössä rajaton: varsinkaan kenttätyön alkuvaiheissa ei välttämättä osaa arvioida, mitä kentältä kannattaa kerätä ja mitä ei tarvitsisikaan. Siksi etnografinen aineisto ja sen käsittely on jonkin verran erilaista kuin sellainen aineisto, jonka koko ja koostumus on ollut tiedossa jo keruun alkaessa (vrt. Korpusaineistot tk.). Aineistoa kertyy yleensä paljon, ja tallenteita esimerkiksi tuskin on mielekästä alkaa litteroida alusta loppuun: tutkimuskysymysten tarkennuttua ja useiden kuuntelukertojen jälkeen voi valita tarkempaan tarkkailuun ne episodit, joiden tarkempi analyysi erityisesti vastaa tutkimuskysymyksiin. Valinnassa auttavat myös vinjetit (Creese, Takhi, & Blackledge 2016), joiden avulla tutkija erottaa olennaisen kohinasta.

## 3.2. Tutkijanpositio ja etiikka

Etnografian eettisyys on merkittävä mutta vaikeasti arvioitava tekijä. Eettinen tavoite kaikessa tutkimuksessa on se, että se ei vahingoita tutkimukseen osallistuvia. Etnografisessa tutkimuksessa yksityisyyden suoja ja tutkimukseen osallistuvan itsemääräämisoikeus ovat keskeisiä eettisiä näkökohtia. Kirjallinen tai suullinen tutkimukseen suostuminen ei vielä tee tutkimuksesta eettistä. Pitkäaikaisessa etnografisessa tutkimuksessa onkin olennaista, että tutkimuksesta puhutaan säännöllisesti, ja tutkimukseen osallistujalla on esimerkiksi mahdollisuus kieltää jonkin tilanteen ääni- tai kuvatallentaminen tai jonkin keskustelun päätyminen tutkimusaineistoksi edes muistiinpanojen muodossa. On myös tutkijan velvollisuus muistuttaa tutkimukseen osallistujia siitä, että tutkija on työssään. Näin on erityisesti laita silloin, kun tutkijan ja tutkimukseen osallistujan suhde kehittyy läheiseksi vuorovaikutussuhteeksi. Tämä on pitkäaikaisessa etnografiassa hyvin todennäköistä.

Etnografisen tutkimusprosessin vuorovaikutteisuus ja syklisyys edellyttävät tutkijalta tietoisuutta omasta roolistaan. Tutkijan tulee aina ottaa huomioon, että tutkimukseen osallistujat ovat omaa toimintaansa analysoivia subjekteja, jotka voivat kriittisesti arvioida tutkijan etnografista kirjoitusta ja hänen eettisiä ja poliittisia näkemyksiään. Tutkijan on hyvä pyrkiä läheiseen vuorovaikutukseen tutkimukseen osallistuvien kanssa ja samaistumaan yhteisön jäsenten kokemusmaailmaan – niissä rajoissa kuin se on mahdollista. Tutkijalla voi olla myös toimintaa muuttava rooli, jolloin ilmiön ymmärryksestä saattaa nousta kriittinen ja kantaa ottava tulkinta. Tästäkin positiosta tutkija neuvottelee toimintayhteisön jäsenten kanssa tutkimuksen aikana.

Tutkijan rooliin liittyy myös ristiriitoja. Esimerkiksi kouluetnografiassa tutkijalle saatetaan kentällä sovittaa stereotyyppistä arvioijan pukua, jossa tutkijan oletetaan arvioivan koulun toimintatapoja etsiäkseen niistä puutteita tai heikkouksia (ks. esim. Pöyhönen 2004). Tämä vaikeuttaa olennaisesti tutkijan mahdollisuuksia ymmärtää koulun toimintaa sisältäpäin ja jopa estää koko tutkimuksen toteutumisen. Oppilaiden arkea koulussa tutkiva aikuinen luovii myös oppilaiden luottamuksen ja aikuisen auktoriteetin välillä: hänellä on aikuisena ja ulkopuolisena

oletusarvoisesti sellaisia oikeuksia, joita oppilailla ei ole, eivätkä säännöt ja velvollisuudet koske häntä niin kuin oppilaita. Tämä tekee tutkijasta oppilaiden silmissä helposti opettajien kumppanin. (Ks. Lehtonen 2015.) Tutkijan on siis tunnistettava ja tehtävä selväksi oman toimintansa ammatilliset rajat. Vastaanottokeskuksen etnografiassa (Pöyhönen, Kokkonen & Tarnanen 2019) tutkimusryhmä teki kovasti työtä, jotta tutkijoita ei pidettäisi maahanmuuttoviraston tai poliisin edustajina. Turvapaikanhakijat olivat oppineet olemaan luottamatta viranomaiseen, ja tutkija asemoitui viranomaiseksi heidän lähtömaissaan. Tätä luottamusta piti rakentaa koko tutkimusprosessin ajan.

Tutkija joutuu kenttätyön kuluessa usein miettimään, miksi tutkittavat suostuisivat hänen pyyntöihinsä; mitä he tutkimuksesta ”saavat”. Ideaali tilanne on tietenkin sellainen, että tutkimukseen osallistuvat kokevat pystyvänsä vaikuttamaan heitä koskeviin ja heille tärkeisiin asioihin ja haluavat siksi osallistua tutkimukseen. Kun toimitaan lasten tai nuorten kanssa, hyvin abstrakteja perusteluja voi olla turha antaa. Heistäkin kuitenkin yleensä on hauskaa saada katsoa tai kuunnella omia tallenteitaan (jos tutkija lupaa tämän, lupaus on pidettävä!). Joskus tutkijat ovat korvanneet tutkittavien heille antamaa aikaa esimerkiksi tarjoamalla kahvit, ruoan tai vaikkapa elokuvalippuja. Osallistuminen pelkän palkkion vuoksi on kuitenkin eettisesti ongelmallista. Tutkijan on pohdittava myös omia voimavarojaan, jos tutkimukseen osallistuja ajattelee saavansa tutkijasta elinikäisen ystävän. Etenkin pitkäaikaisessa etnografisessa tutkimuksessa on hyvä hiljalleen valmistautua tutkimuksen päättymiseen – niin tutkijan kuin tutkimukseen osallistujienkin.

Etnografinen tutkimus ei pääty raportointiin, vaan tutkijan tulee ottaa myös huomioon, millainen ilmapiiri yhteisöön jää tutkimuksen jälkeen ja miten tutkimus mahdollisesti muuttaa yhteisön toimintaa. Tutkijan on myös huolehdittava siitä, että hän ei omalla toiminnallaan estä toisten tutkijoiden pääsyä kentälle. Tässä mielessä tutkijan eettinen vastuu tutkimukseen osallistujien hyvinvoinnista jatkuu vielä kentällä olon ja raportoinnin jälkeen.

## 3.3. Millaista tietoa etnografia tuottaa?

Etnografia on olemukseltaan syvästi laadullista. Se ei pyri vastaamaan sellaisiin kysymyksiin, joihin vastataan tilastollisella merkitsevyydellä. Sen sijaan etnografia voi valaista ilmiöitä tilastojen takana: se lisää ymmärrystä siitä, mikä kaikki arjessa on merkittävää. Tilastollisesti voidaan esimerkiksi havaita eroja siinä, miten erilaiset oppilaat menestyvät koulussa ja sijoittuvat jatko-opintoihin. Tilastotieteellinen analyysi kuitenkin harvoin tarjoaa vastauksia siihen, mitä ovat ne arjen toiminnot, joihin osa oppilaista pystyy osallistumaan toisia paremmin. Etnografiassa on ripaus löytöretkeilyä: etnografisin metodein on mahdollista ymmärtää jonkin sellaisen arkisen seikan merkityksellisyys, joka on silmiemme edessä mutta jota emme tule ajatelleeksi. Indiana Jonesin saappaisiin tutkijan ei silti pidä astua, vaan tutkimusta tehdään yhteistyössä, vaikka tutkijalla on tärkeä rooli ilmiön ymmärtämisessä ja kuvaamisessa tiedeyhteisölle.

Yleistettävyys ja toistettavuus ovat klassisia tieteelliselle tutkimukselle asetettuja vaatimuksia. Näistä kumpikaan ei toteudu etnografiassa aivan klassisimmalla tavalla: kun lähtökohta on se, että tilanteisesta ja paikallisesta kontekstista ei voi irrottaa tai jättää huomiotta mitään muuttamatta sitä, on tietysti mahdotonta ajatella, että jokin tutkimus toistettaisiin eri ajassa ja paikassa juuri samanlaisena. Tämä ei kuitenkaan tarkoita, että etnografiassa tingittäisiin tieteellisyydestä tai että sen tuottama tieto ei olisi luotettavaa. Etnografiassa tieteellisyyden vaatimus liittyy ennen muuta läpinäkyvyyteen ja tutkijanposition tiedostamiseen: tutkijan on kuvattava, miten aineisto on syntynyt ja miten argumentti rakentuu, jotta muut voivat arvioida tuloksen relevanssia ja suhdetta omaan työhönsä. Etnografian yleistettävyys ja toistettavuus taas syntyy siitä, miten tutkija keskusteluttaa havaintojaan ja analyysiaan muun tutkimustiedon kanssa. Etnografi kuvaa tendenssejä ja mekanismeja sekä paikallisten käytänteiden suhdetta yhteiskuntajärjestelmiin. Tutkimuksen kulun kuvauksen ja argumentoinnin on oltava niin läpinäkyvää, että toinen tutkija pystyy arvioimaan sen luotettavuutta. Suhteuttaminen voi olla myös monipaikkaista ja kansainvälistä. Esimerkki monipaikkaisesta ja kansainvälisestä etnografiasta on tutkimus, jossa verrattiin, miten

eurooppalaisen viitekehyksen määritelmää tekstitaidoista on tulkittu Tanskassa ja Suomessa aikuisten maahanmuuttajien kotoutumiskoulutuksessa. (Holm & Pöyhönen 2012.)

Hahmotellessaan lingvistiikan ja etnografian suhdetta Rampton (2006, 394–395) korostaa, että lingvistiikka voi napakoittaa etnografiaa. Lingvistiikassa on vankka puheen tallentamisen, käsittelyn ja analysoinnin perinne, joka etnografiseen tietoon yhdistettynä tarjoaa järeitä työkaluja. Sosiologisesti painottuneet etnografiat saattavat tiedostaa kielen keskeisyyden sosiaalisten merkitysten tuottamisessa, mutta ilman kielitieteellistä käsitteistöä kieltä on vaikea eritellä analyyttisesti. (Sosio)lingvistinen koulutus taas opettaa analysoimaan kielellisten merkitysten syntymisen mekanismeja. Toisaalta Rampton (mp.) tuo esiin, että etnografia voi auttaa lingvistejä tarkastelemaan tutkimuksensa lähtökohtia.

Vaikka lingvistisen etnografian yhtenä vahvuutena voi nähdä yksityiskohtaisen ääni- ja videotallenteiden analysoinnin, muunkinlaisilla aineistoilla on lingvistisessä etnografiassa tärkeä rooli. Copland ja Creese (2015, 52) huomauttavat, että on myös sellaisia lingvistisesti orientoituneita etnografisia tutkimuksia, joiden aineistoissa ei ole lainkaan vuorovaikutustallenteita. Muistiinpanot ja kenttäpäiväkirja auttavat kohdistamaan katsetta toistuviin, merkittäviin tapahtumiin ja toimintaan. Tarkoista muistiinpanoista pystyy myös tietyin reunaehdoin analysoimaan vuorovaikutuskuvioita (ks. Lehtonen 2015). On tietysti otettava huomioon, että tallenteen puuttuessa analyysi perustuu vain hetkellisiin havaintoihin. Jos samantapaisia toimintoja toistuu tilanteissa, havainnot kuitenkin tukevat toisiaan. Olennaista tällaisessa lähestymistavassa on se, että tutkija on havainnoinut vuorovaikutustilanteita riittävän kauan erottaakseen toistuvat ja poikkeukselliset tilanteet toisistaan.

Koska lingvistisen etnografian kysymyksenasettelut usein ohjaavat sosiaalisen indeksisyyden tarkasteluun, tarvitaan niin sanottua metapragmaattista aineistoa: toimintaa, joka kommentoi toimintaa, puhetta, joka kommentoi puhetta. Tätä sisältyy myös spontaaneihin ääni- ja videotallenteisiin mutta usein myös haastatteluihin. Erityisesti metapragmaattista aineistoa pyritään keräämään niin sanotuissa retrospektiivisissä haastatteluissa, joissa osallistujat mahdollisesti aineistoja nähtyään kommentoivat omaa toimintaansa. On kuitenkin keskeistä

ymmärtää, että haastatteluraporttikaan ei ole mikään absoluuttinen totuus tai lopullinen varmistus: sekin on yksi vuorovaikutustilanne muiden joukossa, ja sitä on tarkasteltava sen omilla ehdoilla. Kyse ei siis ole pelkästään aineiston validoinnista, kun tutkimukseen osallistujaa haastatellaan ja tämä reflektoi toimintaansa jossakin vuorovaikutustilanteessa. Kieli, puhe ja toiminta ja näitä kommentoiva kieli, puhe ja toiminta ovat yhtä ja samaa prosessia, jossa merkitykset ja tulkinnat syntyvät. Myös tutkijan toiminta on osa tuota samaa prosessia, minkä hyväksyminen tavallaan mitätöi tarkkailijan paradoksin (paradoksista ks. Haastattelu aineistonkeruun metodina sosiolingvistiikassa tk.). Me kaikki paitsi tuotamme kieltä myös tulkitsemme sitä sekä asemoimme itseämme ja muita tulkintoihimme.

Lingvistisen etnografian aineisto ja analyysi ylittävät mikro- ja makrotason rajan: tavoitteena on juuri kuvata linkki yksilöiden vuorovaikutuksen ja laajempien yhteiskunnallisten valtasuhteiden ja rakenteiden välillä (Copland & Creese 2015). Jotkut tutkijat ovat myös kokonaan kyseenalaistaneet jaon mikro- ja makrotasoon: missä muualla laajemmin jaetut merkityksetkään syntyisivät kuin yksilöiden vuorovaikutustoiminnassa? (Heller 2011; Martin-Jones, Andrews & Martin 2017; ks. myös Diskurssianalyysi tk.) Keskustelunanalyysissa, jolla on yhteistä historiaa lingvistisen etnografian kanssa, on keskitytty nimenomaan vuorovaikutuksen mikrotasoon. Yksi sen lähtökohdista on, että yhteyksiä laajempiin merkityksiin ei pitäisi etsiä tarkasteltavana olevan keskustelukontekstin ulkopuolelta (ks. Keskustelunanalyysi ja vuorovaikutuslingvistiikka tk.). Eri aineistojen ja kontekstien yhdistäminen voikin aiheuttaa metodologisia ristiriitoja lingvistisen etnografian ja keskustelunanalyyttisen perinteen välillä, sillä lingvistinen etnografia ja vuorovaikutussosiolingvistiikka hyödyntävät vuorovaikutuksen analyysissaan keskustelunanalyysin työkaluja mutta eivät sitoudu kaikkiin sen reunaehtoihin, esimerkiksi vaatimukseen siitä, että kaikki analyysin kannalta olennainen olisi osoitettava juuri analysoinnin kohteena olevasta keskusteluaineistosta (ks. esim. Day & Kjaerbeck 2013, 19; Lehtonen 2015, 57–58). Vuorovaikutusanalyysin yhdistäminen laajemmin jaettuihin sosiaalisiin merkityksiin, esimerkiksi se, millaisiin ideologioihin ja stereotypioihin aiemmin mainitut *mina ei suami puhu* -performanssit

liittyvät, onkin kysymys, joka vaatii lingvistisen etnografian harjoittajalta metodologista kurinalaisuutta ja valmiutta kyseenalaistaa omaa analyysiaan ja positiotaan – ja kuitenkin se on sellainen kielen kuvauksen kysymys, johon vastaamisessa juuri lingvistisessä etnografiassa on päästy huomattavan pitkälle.

Lingvistinen etnografia (tai etnografia yleensä) ylittää usein rajoja myös siinä, millä tavalla tutkimustuloksia esitellään ja levitetään. Monissa etnografisissa projekteissa hyödynnetään esimerkiksi myös taiteen keinoja (ks. esim. Haapakangas, Pöyhönen & Serkamo 2017; Bradley & McKay 2016).[5] Taide voi olla analyysin kohde, ja taiteilijoiden ja tutkijoiden työt voivat myös ruokkia toisiaan (Lehtonen & Pöyhönen 2019). Etnografi keskittyy tavallisesti jäljittämään toimintayhteisön vuorovaikutusta, kun taas taiteilijalle voi olla olennaisempaa lopputulos ja toiminnan reflektointi, kuten maalaus tai teatteriesitys ja siitä saatu palaute. Kummatkin tarvitsevat toisiaan tavoitteiden toteuttamiseksi. Joka tapauksessa taiteella ja etnografialla on yhteinen tavoite siinä, miten ne pyrkivät välittämään psykofyysisiä kokemuksia jostakin tilasta, ajasta ja paikasta. Koska etnografian tarkoitus on kuvata osallistujien kokemusta ja näkökulmaa, etnografi on myös kentälle velkaa sen, että tuo tutkittavien äänen kuuluviin yhteyksissä, joissa se ei muutoin kuuluisi mutta joissa heitä koskevista asioista päätetään tai keskustellaan. Siksi yleistajuinen kirjoittaminen tai puhuminen on olennainen osa monen etnografin työtä. Epäoikeudenmukaisten valta-asetelmien tarkkailu saattaa kutsua myös kansalaisaktivismiin, mikä ravistelee tutkijan rooleja ja rajoja (ks. Lehtonen & Pöyhönen 2019).

## 4. Yhteenveto

Olemme tässä artikkelissa kuvanneet lingvistisen etnografian historiallista taustaa ja ymmärrystä kielestä, aineistonkeruun periaatteita ja käytänteitä, tutkijanpositiota ja tutkimusetiikkaa sekä sitä, millaiseen tutkimukseen lingvistinen etnografia soveltuu.

Keskeistä lingvistisessä etnografiassa on tutkimuksen suuntaaminen vuorovaikutukseen toimintayhteisössä ja siihen, mitä merkitystä tällä vuorovaikutuksella on toimintaan osallistuville yksilöille ja laajemmin yhteiskuntajärjestelmälle. Olennaista on tunnistaa kielen ja vuorovaikutuksen kompleksisuus rajaamatta sitä muutamiin muuttujiin tai tietyssä tilanteessa tapahtuvaan toimintaan. Lingvistinen etnografia avaa siten ikkunan myös ajallis-paikalliseen ulottuvuuteen ja ottaa huomioon vuorovaikutuksen toistuvuudet ja poikkeavuudet sekä pysyvyyden ja muutoksen.

Lingvistinen etnografia, kuten etnografia laajemminkin, on tutkimuseettisesti haastavaa. Se edellyttää ymmärrystä siitä, että tärkeintä etnografisessa tutkimuksessa on luottamuksen rakentaminen ja säilyttäminen tutkijan ja tutkimukseen osallistuvien välillä. Luottamuksen voi menettää varomattomalla toiminnalla kentällä. Huolellinen valmistelu ja oman toiminnan analyyttinen tarkastelu auttavat tutkijaa tässä prosessissa.

**Aiheesta lisää:**

Blommaert, Jan & Jie, Dong. 2010. *Ethnographic Fieldwork. A Beginner's guide*. Bristol: Multilingual Matters.

Copland, Fiona & Creese, Angela. 2015. *Linguistic Ethnography: Collecting, Analysing and Presenting Data*. London: Sage.

Gardner, Sheena & Martin-Jones, Marilyn (toim.). 2012. *Multilingualism, Discourse and Ethnography.* London: Routledge.

Heller, Monica, Pietikäinen, Sari & Pujolar, Joan. 2018. *Critical Sociolinguistic Research Methods. Studying Language Issues that Matter.* London: Routledge.

## VIITTEET

1 *Sosiaalinen semioosi* käsitteenä liittyy Peircen (ks. esim. 1955) semiotiikkaan, jota sittemmin on hyödynnetty indeksisyyttä ja kielen sosiaalista merkitystä tarkastelevassa lingvistisessä antropologiassa (Silverstein 2003; Agha 2007; Nieminen 2010; Visakko 2015).

2 Esimerkkinä variaabelista on yksikön 1. persoonan pronomini, jonka variantteja aineistossa voivat olla vaikkapa *minä, mä* ja *mie*.

3 *Lekti*-termin lähin suomennos lienee juuri kielimuoto. Lektiä on käytetty kuvaamaan mitä tahansa melko selvärajaista kielellisten merkkien joukkoa, jota tietynlaiset ihmiset käyttävät tiettyihin tarkoituksiin. Usein lekti on jotenkin nimettävissä, ja lektien rajaus voi olla esimerkiksi alueellista (engl. *dialect*), sosiaalista (engl. *sosiolect*) tai jopa puhujakohtaista – tällöin on puhuttu *idiolektistä*. Quist (2008) käsittelee artikkelissaan kahdenlaista lähestymistapaa sosiolingvistiseen variaatioon: sellaista, joka ottaa tarkastelun pohjaksi kielimuodon, ja sellaista, joka näkee kielen sosiaalisena toimintana.

4 ”The total linguistic fact, the datum for science of language is irreducibly dialectic in nature. It is an unstable mutual interaction of meaningful sign forms, contextualised to situations of interested human use and mediated by the fact of cultural ideology.” (Silverstein mp.)

5 Meneillään olevista hankkeista voidaan mainita Koneen Säätiön rahoittamat Liisa Raevaaran johtama Taide työnä ja työvälineenä (http://urbduuni.fi/), Jyrki Kalliokosken johtama Itä-Helsingin uudet Suomen kielet (http://blogs.helsinki.fi/uudetsuomet/) sekä Jussi Lehtosen ja Sari Pöyhösen Toinen koti -dokumenttiteatteriprojekti (http://www.kiertuenayttamo.fi/esitykset/toinen-koti/).

## KIRJALLISUUS

Agha, Asif. 2007. *Language and Social Relations*. Cambridge: Cambridge University Press.

Anderson, Benedict. 1983. *Imagined Communities. Reflections on the Origins and Spread of Nationalism*. London: Verso.

Androutsopoulos, Jannis (toim.). 2014. *Mediatization and Sociolinguistic Change*. Berlin: de Gruyter.

Atkinson, Paul, Delamont, Sara & Housley, William. 2008. *Contours of Culture*: *Complex Ethnography and the Ethnography of Complexity*. Plymouth: Altamira Press.

Atkinson, Paul & Hammersley, Martyn. 1998. Ethnography and participant observation. Julkaisussa: Denzin, Norman K. & Lincoln, Yvonna S. (toim.) *Strategies of Qualitative Inquiry*. London: Sage, 111–136.

Auer, Peter. 1992. Introduction: John Gumperz' approach to contextualization. Julkaisussa: Auer, Peter & Di Luzio, Aldo (toim.) *The Contextualization of Language*. Amsterdam: John Benjamins, 1–38.

Auer, Peter (toim.). 2007. *Style and Social Identities. Alternative Approaches to Linguistic Heterogeneity*. Language, Power and Social Processes 18. Berlin: Mouton de Gruyter.

Barton, David & Tusting, Karin (toim.). 2005. *Beyond Communities of Practice. Language, Power and Social Context.* Cambridge: Cambridge University Press.
Blommaert, Jan. 2010. *The Sociolinguistics of Globalization*. Cambridge: Cambridge University Press.
Blommaert, Jan & Jie, Dong. 2010. *Ethnographic Fieldwork. A Beginner's Guide.* Bristol: Multilingual Matters.
Blommaert, Jan & Rampton, Ben. 2011. Language and superdiversity. *Diversities* 13:2, 1–21. Unesco.
Bradley, Jessica & McKay, Sam. 2016. How does arts practice engage with narratives of migration from refugees? Lessons from 'utopia'. *Journal of Arts & Communities* 8:1–2, 31–46.
Bradley, Jessica & Moore, Emilee & Simpson, James & Atkinson, Louise. 2018. Translanguaging space and creative activity: Collaborative ethnography and arts-based learning. *Language and Intercultural Communication* 18:1, 54–73.
Bäckman, Linda. 2017. *'Second Generation?' Language and Identity among Adults whose Parents were Migrants*. Turku: Åbo Akademi.
Copland, Fiona & Creese, Angela. 2015. *Linguistic Ethnography: Collecting, Analysing and Presenting Data*. London: Sage.
Coupland, Nikolas. 2007. *Style. Language Variation and Identity*. Cambridge: Cambridge University Press.
Creese, Angela, Takhi, Jaspreet Kaur & Blackledge, Adrian. 2016. Reflexivity in team ethnography: Using researcher vignettes. Julkaisussa: Martin-Jones, Marilyn & Martin, Deirdre (toim.) *Researching Multilingualism. Critical and Ethnographic Perspectives*. London: Routledge, 189–214.
Day, Dennis & Kjaerbeck, Susanne. 2013. 'Positioning' in the conversation analytic approach. *Narrative Inquiry* 23:1, 16–39.
Deegan, Mary Jo. 2007 [2001]. The Chicago School of Ethnography. Julkaisussa: Atkinson, Paul, Coffey, Amanda, Delamont, Sara, Lofland, John & Lofland, Lyn (toim.) *Handbook of Ethnography*. London: Sage, 11–25.
Duranti, Alessandro. 1997. *Linguistic Anthropology*. Cambridge: Cambridge University Press.
Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12:4, 53–76.
Eckert, Penelope. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41, 87–100.
Eckert, Penelope & McConnell-Ginet, Sally. 1992. Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology* 21, 461–490.
Emerson, Robert M. & Fretz, Rachel I. & Shaw, Linda L. 2007 [2001]. Participant observation and field notes. Julkaisussa: Atkinson, Paul, Coffey, Amanda, Delamont, Sara, Lofland, John & Lofland, Lyn (toim.) *Handbook of Ethnography*. London: Sage, 352–369.
Geertz, Clifford. 1973. *The Interpretation of Cultures. Selected Essays*. New York (NY): Basic Books.
Gordon, Tuula, Hynninen, Pirkko, Lahelma, Elina, Metso, Tuija, Palmu, Tarja & Tolonen, Tarja. 2007. Koulun arkea tutkimassa. Kokemuksia kollektiivisesta etnografiasta.

Julkaisussa: Lappalainen, Sirpa, Hynninen, Pirkko, Kankkunen, Tarja, Lahelma, Elina & Tolonen, Tarja (toim.) *Etnografia metodologiana. Lähtökohtana koulutuksen tutkimus.* Tampere: Vastapaino, 41–65.

Guba, Egon G. & Lincoln, Yvonna S. 1998. Competing paradigms in qualitative research. Julkaisussa: Denzin, Norman K. & Lincoln, Yvonna S. (toim.) *The Landscape of Qualitative Research. Theories and Issues*. London: Sage, 195–220.

Gumperz, John J. 1982. *Discourse Strategies*. Studies in Interactional Sociolinguistics 1. Cambridge: Cambridge University Press.

Gumperz, John J. 1992. Contextualization revisited. Julkaisussa: Auer, Peter & Di Luzio, Aldo (toim.) *The Contextualization of Language*. Amsterdam: John Benjamins, 39–54.

Haapakangas, Eeva-Leena, Pöyhönen, Sari & Serkamo, Pia. 2017. "Toinen koti" rajojen ylittäjänä. *Kieli, koulutus ja yhteiskunta* 8:4. [verkkoaineisto]. [viitattu 16.3.2018]. Saatavissa: https://www.kieliverkosto.fi/fi/article/toinen-koti-rajojen-ylittajana.

Hammersley, Martyn. 2007. Reflections on linguistic ethnography. *Journal of Sociolinguistics* 11:5, 689–695.

Heller, Monica. 2011. *Paths to Post-nationalism. A Critical Ethnography of Language and Identity*. Oxford: Oxford University Press.

Heller, Monica, Pietikäinen, Sari & Pujolar, Joan. 2018. *Critical Sociolinguistic Research Methods. Studying Language Issues that Matter*. London: Routledge.

Holm, Lars & Pöyhönen, Sari. 2012. Localising supranational concepts of literacy in adult second language teaching. Julkaisussa: Pitkänen-Huhta, Anne & Holm, Lars (toim.) *Literacy practices in transition: Perspectives from the Nordic countries*. Bristol: Multilingual Matters, 191–210.

Hymes, Dell. 1972a. On communicative competence. Julkaisussa: Pride, John B. & Holmes, Janet (toim.) *Sociolinguistics: Selected Readings*. Harmondsworth: Penguin Books, 269–293.

Hymes, Dell. 1972b. Models of the interaction of language and social life. Julkaisussa: Gumperz, John & Hymes, Dell (toim.) *Directions in Sociolinguistics. The Ethnography of Communication*. New York (NY): Holt, Rinehart & Winston, 35–71.

Hymes, Dell. 1974. *Foundations in Sociolinguistics. An Ethnographic Approach*. Philadelphia: University of Pennsylvania Press.

Hymes, Dell & Gumperz, John (toim.). 1972. *Directions in Sociolinguistics. The Ethnography of Communication*. New York (NY): Holt, Rinehart & Winston.

Jørgensen, Jens Normann, Karrebæk, Martha Siv, Madsen, Lian Malai & Møller, Janus Spindler. 2011. Polylanguaging in superdiversity. *Diversities* 13:2, 23–37.

Jørgensen, Jens Normann & Møller, Janus Spindler. 2014. Polylingualism and languaging. Julkaisussa: Leung, Constant & Street, Brian V. (toim.) *The Routledge Companion to English Studies*. New York (NY): Routledge, 67–83.

Kytölä, Samu. 2013. *Multilingual Language Use and Metapragmatic Reflexivity in Finnish Internet Football Forums. A Study in the Sociolinguistics of Globalization*. Jyväskylä Studies in Humanities 200. Jyväskylä: Jyväskylän yliopisto.

Lappalainen, Sirpa, Hynninen, Pirkko, Kankkunen, Tarja, Lahelma, Elina & Tolonen, Tarja (toim.). 2007. *Etnografia metodologiana. Lähtökohtana koulutuksen tutkimus*. Tampere: Vastapaino.

Lave, Jean & Wenger, Etienne. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.

Lehtonen, Heini. 2015. *Tyylitellen. Nuorten kielelliset resurssit ja kielen sosiaalinen indeksisyys monietnisessä Helsingissä*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-1333-7.

Lehtonen, Heini. 2016. Troping on prejudice. *Stylised "bad Finnish"* performances and reflexivity among adolescents in Eastern Helsinki. *AILA Review* 29, 15–47.

Lehtonen, Jussi & Pöyhönen, Sari. 2019. Documentary theatre as a platform for hope and social justice. Julkaisussa: Anttila, Eeva & Suominen, Anniina (toim.) *Critical Articulations of Hope from the Margins of Arts Education. International Perspectives and Practices*. London: Routledge, 31–44.

Madsen, Lian Malai & Karrebæk, Martha Sif & Møller, Janus Spindler (toim.). 2015. *Everyday Languaging: Collaborative Research on the Language Use of Children and Youth. Trends in Applied Linguistics*. Berlin: Mouton de Gruyter.

Marcus, George E. 1995. Ethnography in/of the world system: The emergence of multi-sited ethnography. *Annual Review of Anthropology* 24, 95–117.

Martin-Jones, Marilyn, Andrews, Jane & Martin, Deirdre. 2017. Reflexive ethnographic research practice in multilingual contexts. Julkaisussa: Martin-Jones, Marilyn & Martin, Deirdre (toim.) *Researching Multilingualism. Critical and Ethnographic Perspectives*. London: Routledge, 189–202.

Meyerhoff, Miriam. 2002. Communities of practise. Julkaisussa: Chambers, John K., Trudgill, Peter & Schilling-Estes, Natalie (toim.) *The Handbook of Language Variation and Change*. Oxford: Blackwell, 526–548.

Mononen, Kaarina. 2013. *Inkerinsuomalaisten suomen kielen käyttö Pietarissa ja sen lähialueella*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-8657-1.

Nieminen, Tommi. 2010. *Lajien synty: tekstilaji kielitieteen semioottisessa metateoriassa*. Väitöskirja. Jyväskylä: Jyväskylän yliopisto.

Ochs, Elinor. 1992. Indexing gender. Julkaisussa: Duranti, Alessandro & Goodwin, Charles (toim.) *Rethinking Context: Language as an Interactive Phenomenon*. Cambridge: Cambridge University Press, 335–358.

Peirce, Charles. 1955. *Philosophical Writings of Peirce*. Toimittanut Justus Buchler. New York (NY): Dover Publications.

Peuronen, Saija. 2017. *Heteroglossic and Multimodal Resources in Use: Participation across Spaces of Identification in a Christian Lifestyle Sports Community*. Jyväskylä Studies in Humanities 305. Jyväskylä: Jyväskylän yliopisto.

Pöyhönen, Sari. 2004. *Suomen kielen opettajana Venäjällä. Ammatti-identiteetin tulkintoja koulutuksen ja opetuksen murroksessa*. Jyväskylä: Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus.

Pöyhönen, Sari, Kokkonen, Lotta & Tarnanen, Mirja. 2019. Turvapaikanhakijoiden kertomuksia sosiaalisista verkostoista ja kuulumisen tunteista. Julkaisussa: Lyytinen, Eveliina (toim.) *Turvapaikanhaku ja pakolaisuus Suomessa*. Turku: Siirtolaisuusinstituutti, 183–204. Saatavissa: https://siirtolaisuusinstituutti.fi/wp-content/uploads/2019/12/t-02-isbn-978-952-7167-60-1-turvapaikanhaku-ja-pakolaisuus-suomessa.pdf.

Quist, Pia. 2005. *Stilistiske praksisser i storbyens heterogene skole – en etnografisk og sociolinguistisk undersøkelse af sproglig variation.* Nordisk Forskningsinstitut, Afdelning for dialektforskning. København: Københavns Universitet.

Quist, Pia. 2008. Sociolinguistic approaches to multiethnolect: Language variety and stylistic practice. *International Journal of Bilingualism* 12:1–2, 43–61.

Rampton, Ben. 2006. *Language in Late Modernity. Interaction in an Urban School.* Cambridge: Cambridge University Press.

Rampton, Ben. 2007. Neo-Hymesian linguistic ethnography in the United Kingdom. *Journal of Sociolinguistics* 11:5, 584–607.

Rampton, Ben, Maybin, Janet & Roberts, Celia. 2015. Theory and method in linguistic ethnography. Julkaisussa: Snell, Julia, Shaw, Sarah & Copland, Fiona (toim.) *Linguistic Ethnography. Interdisciplinary Explorations.* London: Palgrave, 14–50.

Rampton, Ben, Tustin, Karin, Maybin, Janet, Barwell, Rickhard, Creese, Angela & Lytra, Vally. 2004. *UK Linguistic ethnography: a discussion paper.* Coordinating Committee, UK Linguistic Ethnography Forum. [verkkoaineisto]. [viitattu 16.3.2018]. Saatavissa: http://www.lancs.ac.uk/fss/organisations/lingethn/documents/discussion_paper_jan_05.pdf.

Ruohonen, Ilkka & Laitila, Teuvo. 1994. Tulkintojen ja symbolien ytimessä: geertziläinen näkökulma kulttuureihin. Julkaisussa: Ruohonen, Ilkka & Laitila, Teuvo (toim.) *Näköaloja kulttuureihin: antropologian historiaa ja nykysuuntauksia*. Helsinki: Gaudeamus, 116–135.

Räisänen, Tiina. 2013. *Professional Communicative Repertoires and Trajectories of Socialization into Global Working Life*. Jyväskylä Studies in Humanities, 216. Jyväskylä: Jyväskylän yliopisto.

Sherman Heyl, Barbara. 2007 [2001]. Ethnographic interviewing. Julkaisussa: Atkinson, Paul, Coffey, Amanda, Delamont, Sara, Lofland, John & Lofland, Lyn (toim.) *Handbook of Ethnography.* London: Sage, 369–383.

Silverstein, Michael. 1985. On the pragmatic "poetry" of prose. Julkaisussa: Deborah Schiffrin (toim.) *Meaning, Form and Use in Context*. Washington (DC): Georgetown University Press, 181–199.

Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language and Communication* 23, 193–229.

Vaattovaara, Johanna & Halonen, Mia. 2015. Missä on ässä? "Stadilaisen s:n" helsinkiläisyydestä. Julkaisussa: Sorjonen, Marja-Leena, Rouhikoski, Anu & Lehtonen, Heini (toim.) *Helsingissä puhuttavat suomet. Kielen indeksisyys ja sosiaaliset identiteetit.* Helsinki: Suomalaisen Kirjallisuuden Seura, 40–83.

Visakko, Tomi. 2015. *Self-promotion as Semiotic Behavior. The Mediation of Personhood in Light of Finnish Online Dating Advertisements.* Väitöskirja. Helsinki: Helsingin yliopisto.

Wenger, Etienne. 1998. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge: Cambridge University Press.

Westinen, Elina. 2014. *The Discursive Construction of Authenticity: Resources, Scales and Polycentricity in Finnish Hip Hop Culture*. Jyväskylä studies in humanities 227. Jyväskylä: Jyväskylän yliopisto.

Woolard, Kathryn A. 2008. Why *dat* now?: Linguistic-anthropological contributions to the explanation of sociolinguistic icons and change. *Journal of Sociolinguistics* 12:4, 432–452.

# Luku 6 Diskurssianalyysi

*Pekka Pälli*
https://orcid.org/0000-0003-0287-8813

*Ella Lillqvist*
https://orcid.org/0000-0001-7419-4491

**Mikä?**

Diskurssianalyysi on väljä, tieteenalarajat ylittävä tutkimusalue, jota yhdistää kiinnostus siihen, mitä kielellä tehdään ja miten kieli ja sen käyttö yhdistyvät muihin yhteiskunnallisiin ja sosiaalisiin rakenteisiin. Diskurssianalyysi on ennen kaikkea empiirinen, luonnolliseen kielenkäyttöön kohdistuva analyysitapa, jossa ensisijaisena tulkinnan kohteena on se, miten asiat sanotaan, eikä niinkään se, mitä sanotaan. Diskurssianalyysi on riippumaton tulkittavana olevan aineiston kielimuodosta, ja tutkimuskohteena olevaa ilmiötä ymmärtääkseen diskurssianalyytikko voi käyttää aineistonaan erilaisia kielimuotoja samaan aikaan.

**Katso myös:**

Johdannon luku 1.6. Miten kielitieteen menetelmillä voidaan vastata ihmiskunnan suuriin haasteisiin?

Luku 20 Tekstintutkimus

Luku 22 Keskustelunanalyysi ja vuorovaikutuslingvistiikka

# 1. Johdanto

Zellig Harrisin vuonna 1952 uutena analyysimenetelmänä esittelemä diskurssianalyysi perustui siihen ajatukseen, että kielitieteessä – ja kielen kuvauksessa – käytetyt perinteiset kategoriat fonologia (äänneoppi), morfologia (muoto-oppi), leksikologia (sanaoppi) ja syntaksi (lauseoppi) eivät mahdollista lausetta laajempien yksiköiden mielekästä kuvaamista. Niinpä tarvittiin syntaktista tasoa laajempi käsitteellinen kategoria, joksi Harris esitteli diskurssin. Kun siis esimerkiksi sanastontutkijan analyysin kohde oli sanataso ja syntaktikon lausetaso, diskurssin analyysin kohteeksi asettui loogisesti diskurssin taso eli kokonaisuus, joka on laajempi kuin vain lause tai virke. Tässä mielessä diskurssianalyysin historia on vahvasti lingvistinen.

Nykyisin diskurssi käsitetään kuitenkin yleisesti eri tavalla: voisimme määritellä sen yleisellä tasolla sosiaalisen toiminnan ja kielenkäytön yhteenkietoutumiseksi. Diskurssianalyysi on siis sellaista kielenkäytön tutkimusta, jossa kielellinen tulkinta yhdistetään sosiaalisen toiminnan tarkasteluun. Brian Paltridge (2012, 2) ilmaisee tällaisen diskurssianalyysin yleisen tehtävän onnistuneesti todetessaan, että diskurssianalyysin tavoitteena on yhdistää kielen säännönmukaisuuksia sosiaalisen toiminnan ja kontekstin säännönmukaisuuksiin. Ensisijaisena tavoitteena on ymmärtää, mitä kielellä tehdään ja millaiseen sosiaaliseen toimintaan se yhdistyy.

Tässä artikkelissa avaamme diskurssianalyysin taustoja ja niihin liittyviä teoreettis-metodologisia sitoumuksia. Yhdymme diskurssianalyysin yleiseen näkemykseen, jonka mukaan diskurssianalyysi ei ole yksi yhtenäinen metodi vaan paremminkin orientaatio tutkia sosiaalisen todellisuuden ilmiöitä ja sosiaalista toimintaa kielenkäytön kautta (van Dijk 1997). Tarkemmat analyysin metodit täytyy valita esimerkiksi aineistotyypin ja tutkimuksen tavoitteen mukaan. Osin tämän vuoksi onkin viime vuosina yleistynyt *diskurssianalyysin* sijasta käytettävä nimitys *diskurssintutkimus* (engl. *discourse studies*) (esim. Angermuller, Maingueneau & Wodak 2014; Wodak & Meyer 2016 [2001]). Mielestämme moninaisuudesta huolimatta on kuitenkin aiheellista puhua diskurssianalyysista myös omana metodisena otteenaan; se on varsin vakiintunut useilla

tieteenaloilla, ja sillä on myös yhteisiä peruspilareita. Näistä olennaisimmat ovat käsitys siitä, että tulkinnan kohteena on todellinen kielenkäyttö, ja siitä, että kielenkäytön nähdään paitsi heijastavan myös rakentavan sosiaalista todellisuutta (ks. Johdannon luku 2 tässä kirjassa).

Käsittelemme eritoten asioita, joiden katsomme olevan yhteisiä diskurssianalyysin erilaisissa toteutustavoissa. Diskurssianalyysia harjoittavina tutkijoina, opettajina ja aikoinaan myös opiskelijoina saamamme kokemuksen perusteella haluamme painottaa myös diskurssianalyysin tekemisen käytännön näkökulmaa, eli pyrimme antamaan lukijalle käytännön välineitä aineiston analysointiin. Vaikka kaikki esittelemämme työkalut eivät ole kaikille diskurssianalyysin tekijöille tärkeitä tai edes käyttökelpoisia, tavoitteemme on kuitenkin kirjoittaa niin, että lukija saisi metodisia mahdollisuuksia ryhtyä rakentamaan oman tutkimuksensa analyysia.

Tämä artikkeli etenee seuraavasti: Ensin, osassa kaksi, käsittelemme diskurssianalyysin moninaista historiaa, lähtökohtia ja teoriataustaa niin maailmalla kuin Suomessa. Osassa kolme käymme läpi eri suuntausten lähtökohtia sekä mahdollisia aineistoja ja analyysin työkaluja. Analyysin keskeisiä käsitteitä on kootusti selitetty myös taulukossa 1. Sen jälkeen konkretisoimme näiden työkalujen käyttöä kahden pidemmän aineistoesimerkin avulla. Lopussa on yhteenveto sekä suositeltua lisälukemistoa.

**Mihin?**

Diskurssianalyysi voi kohdistua yhtä lailla puhuttuun kuin kirjoitettuun kieleen, tai sen aineisto voi olla näiden kielellisten modaliteettien ohessa tavalla tai toisella visuaalinen. Usein diskurssianalyysi yhdistyy ainakin mielikuvissa laadulliseen analyysiin, mutta yhä yleisempää on myös korpuslingvististen määrällisten menetelmien soveltaminen. Diskurssianalyysi sopii metodiksi erityisesti sellaisiin tutkimusasetelmiin, joissa keskeistä on saada tietoa jonkin ilmiön, asian tai käytännön yhteiskunnallisesta tai kulttuurisesta ymmärryksestä ja kielen roolista tuon ymmärryksen tuottajana.

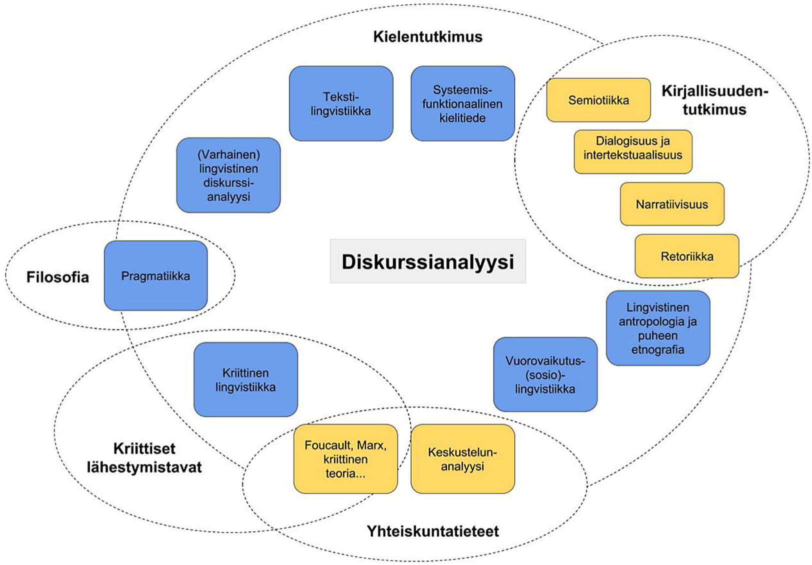
# 2. Diskurssianalyysin historiaa ja taustaa

## 2.1. Diskurssianalyysin alkutaival

Kielentutkimuksellisesti käsite *diskurssianalyysi* voidaan jäljittää Harrisin (1952) *Language*-lehdessä julkaistuun artikkeliin ”Discourse analysis”, jossa Harris ehdotti *diskurssianalyysia* nimitykseksi tietylle metodille, joka tarkastelee puheen tai kirjoituksen kielellisten rakenneyksiköiden yhdistelmää, sidosteista kokonaisuutta (engl. *connected speech/discourse*). Vaikkakin nykyinen ja varsinkin yhteiskuntatieteissä vakiintunut käsitys korostaa *diskurssin* olevan nimenomaan tulkinnallinen, abstrakti käsite, alkuperäinen kielitieteellinen näkemys siis viittasi *diskurssilla* olemassa olevaan kielelliseen kokonaisuuteen. Diskurssianalyysilla ja *diskurssi*-käsitteen ymmärryksellä on kuitenkin useita historiallisia juuria, joita kuvaamme alla (kuva 1). Historiallisten vaikutteiden ja juurten moninaisuus avaa osaltaan ymmärrystä myös nykypäivän diskurssianalyysi(e)n moninaisuuteen.

Harrisin kanssa samoihin aikoihin, jo 1930-luvulta lopulta alkaen, filosofiassa oli vakiinnuttanut jalansijaa niin kutsuttu arkikielen filosofia, jolla oli ja on huomattavia vaikutuksia diskurssianalyyttiseen metodiin. Arkikielen filosofiassa (esim. Wittgenstein 2001 [1953]) painotettiin ajatusta merkityksestä käyttönä, ja kielitieteellisessä metodiikassa tämä painotus sai myöhemmän diskurssianalyysin kannalta vaikutusvaltaisimman muotoilun kielifilosofi J. L. Austinin puheaktiteoriassa (Austin 1962) ja sitä täydentävissä näkemyksissä (esim. Searle 1975). Puheaktiteorian kantava ajatus on se, että kielenkäyttö on toimintaa ja tekoja, eli se on seurauksellista, usein jopa todellisuutta muuttavaa tekemistä. Tämä keskeinen ajatus leimaa myös tämän päivän erilaisia diskurssianalyyttisia tutkimusotteita.

Puheaktiteoriaa voidaan pitää modernin lingvistisen pragmatiikan lähtölaukauksena, ja diskurssianalyysiin se kietoutui erityisesti 1970-luvulta alkaen harjoitetussa erilaisten kielenkäyttötilanteiden, voittopuolisesti nimenomaan keskustelutilanteiden, kuten luokkahuonekeskusteluiden (Sinclair & Coulthard 1975), tutkimuksessa. Nämä diskurssianalyysit siis tyypillisesti sovelsivat pragmatiikkaa, lähinnä




Kuva 1. Diskurssianalyysin vaikutteita. Kuvassa asiaa lähestytään kielentutkimuslähtöisesti niin, että kielentutkimuksen alueelle (iso ympyrä pohjalla) on tullut vaikutteita ulkopuolelta: filosofiasta, yhteiskuntatieteistä ja kirjallisuudentutkimuksesta (keltaisella merkityt alat). Sinisellä merkityt alat ovat ensin kehittyneet osaksi kielitiedettä ja sitten vaikuttaneet myös diskurssianalyysiin. Näiden vaikutteiden lisäksi syvemmällä, tieteenfilosofian tasolla on huomattava sosiaalisen konstruktionismin (esim. Berger & Luckmann 1966) vaikutus.

puheaktiteoriaa, suullisissa vuorovaikutustilanteissa ilmenevien kielellisten tekojen analyysiin.

Diskurssianalyysi, joka 1970-luvulla vähitellen nosti päätään omana nimikkeenään, lainasi metodeja pragmatiikan lisäksi myös sosiolingvistiikan piirissä syntyneistä puheen etnografiasta ja lingvistisestä antropologiasta (Hymes 1974) sekä vuorovaikutussosiolingvistiikasta (Gumperz 1982), joissa oli olennaista eritellä sitä, kuinka tilannekontekstin tekijät vaikuttavat kieleen. Puheen etnografian vaikutukset ovat hyvin selvästi näkyvissä myös 2000-luvun diskurssianalyysissa, jossa etnografiaa hyödynnetään esimerkiksi Hymesin oppeihin perustuen (Blommaert 2009). Yleisemminkin voi sanoa, että diskurssianalyysi laajasti ottaen

jakaa puheen etnografian ja vuorovaikutussosiolingvistiikan keskeisen metodisen kiinnostuksen tilannekontekstin tekijöiden erittelyyn (ks. Lingvistinen etnografia toimintayhteisöissä; Sosiolingvistiikka tk.).

Pragmatiikan ja sosiolingvistiikan lisäksi alun perin sosiologian alueella syntynyt keskustelunanalyysi (ks. Keskustelunanalyysi ja vuorovaikutuslingvistiikka tk.) on ollut diskurssianalyysin kehityksessä tärkeä tutkimuksellinen lähimaasto. Myös diskurssianalyysissa toki jo 1970-luvulla tutkittiin keskusteluja, mutta pääpaino oli melko jähmeästi keskustelussa käytettyjen puheaktien ja keskustelujen kielellisen rakenteen kuvauksessa, kun taas keskustelunanalyysi keskittyi vuorovaikutuksen sosiaalisuuden kuvaukseen. Diskurssianalyysin kehityksen kannalta on joka tapauksessa hyvin olennaista, että vähitellen diskurssianalyysi (tai diskurssintutkimus) ja keskustelunanalyysi ovat monilta osin kietoutuneet toisiinsa. Tätä seikkaa havainnollistaa se, että keskustelunanalyysin käyttäminen tai soveltaminen on yleistä diskurssintutkimuksen valtalehdissä (esim. *Discourse Studies*, *Discourse & Society*, *Discourse and Communication* tai *Text & Talk*) julkaistuissa artikkeleissa.

1970- ja 1980-luvuilla syntyi, ehkä implisiittisestikin, näkemys, että diskurssianalyysi on lähinnä puhutun kielen tutkimusta ja että *diskurssi* merkitsee puhetta tai keskustelua nimenomaan puhuttuna kielenä. Historiallisia jälkiä tästä löytyy vieläkin. Usein kategoriat *teksti* ja *diskurssi* rinnastetaan esimerkiksi opintokokonaisuuksien tai tutkimussuuntausten nimissä siten, että puhutaan tekstin- ja diskurssintutkimuksesta (ks. Tekstintutkimus tk.). Implisiittisesti viitataan jonkinlaiseen jakolinjaan, joka on olemassa erityisesti kielentutkijoiden keskuudessa; sen sijaan yhteiskuntatieteiden parissa tehdyn diskurssianalyysin nimityksiä aineistotyyppi tai kielimuoto (puhuttu tai kirjoitettu) ei näytä ohjaavan samalla tavoin. Varsinkin kansainvälisesti katsoen vaikuttaa kuitenkin siltä, että *diskurssianalyysin* (tai *diskurssintutkimuksen*) käsitteen ala on nykyisin sellainen, että se kattaa kaikenlaisen kielenkäytön.

## 2.2. Kielellisestä rakenteesta kohti yhteiskunnallis-sosiaalisia tulkintoja

Tekstien ja diskurssin tutkimuksen historiaan liittyy kiinnostavia rinnakkaisia kehityskulkuja, joista erityisen mielenkiintoisia nykypäivän ymmärryksen kannalta tekee niiden konvergenssi ja osin sulautuminen yhteen – ainakin siinä mielessä, että erilaiset ja erilaisista traditioista kasvaneet tutkimusotteet ovat tulleet osaksi diskurssintutkimusta. Lyhyesti voi sanoa, että 1970- ja 1980-luvuilla diskurssianalyysi eri muodoissaan oli painottunut kielen merkitysresurssien erittelyyn ja niiden toiminnan logiikan esittelyyn (esim. Brown & Yule 1983; Schiffrin 1994), kun taas sittemmin diskurssianalyysi kehittyi kohti sellaista painotusta, jossa ensisijainen analyysin kohde on yhteiskunta tai sosiaaliset ilmiöt ja jossa kielen merkitysresurssit ovat välineitä näiden ilmiöiden tarkastelussa.

Kehitystä kuvaa hyvin yhden diskurssintutkimuksen merkittävän suunnannäyttäjän, Teun van Dijkin tuotanto. Van Dijk tuli 1970-luvulla tunnetuksi erityisesti tekstien kieliopillisten rakenteiden, eritoten merkitysrakenteiden tutkijana (esim. van Dijk 1972; 1977). Sittemmin, 1980-luvulta alkaen, van Dijk siirtyi tutkimaan painokkaasti yhteiskunnallisia aiheita. 1980-luvun loppupuolella voimistui yleisemminkin se näkemys, että diskurssianalyysissa ei ensisijaisesti ole kyse kielen vaan nimenomaan yhteiskunnan analyysista. Van Dijkin lisäksi merkittävistä diskurssianalyysin suunnannäyttäjistä samansuuntainen kehityskulku, lingvistisen analyysin yhteiskunnallistuminen, näkyy esimerkiksi Norman Faircloughhin ja Ruth Wodakin analyyseissa ja niiden kohdeyleisöissä.

Yhteiskunnallisuus ei toki ollut 1980-luvulla diskurssianalyysiin kehitetty uusi mauste. Jo 1970-luvulla kehittynyt kriittinen lingvistiikka oli painokkaan yhteiskunnallista ja poliittista (esim. Fowler, Hodge, Kress & Trew 1979), ja varsinkin kriittisenä diskurssianalyysina myöhemmin tunnetuksi tullut diskurssianalyysin muoto on monilta osin luontevaa jatkumoa kriittiselle lingvistiikalle. 1970-luvun kriittinen lingvistiikka myös hyödynsi tekstisemanttisia, erityisesti kirjoitettujen tekstien analyysiin kehitettyjä välineitä, joista tunnetuimpana ja eniten

sovellettuna – ja ehkä myös kokonaisvaltaisimpana – voitaneen pitää M. A. K. Hallidayn systeemis-funktionaalista lähestymistapaa kielen kuvaukseen (Halliday 1994 [1985]). Hallidayn lähestymistavan merkityksellisyyttä yhtenä diskurssianalyysin päämetodina on korostanut ennen kaikkea se, että varsinkin Fairclougin (esim. 1992; 2001 [1989]; 2003) hahmottelema kriittisen diskurssianalyysin analyysitapa nojaa vahvasti systeemis-funktionaaliseen kielenkuvauksen malliin.

Nykypäivän diskurssianalyysia on hyvin vaikea nähdä pelkästään kielitieteellisenä metodina, sillä diskurssianalyyttisen metodiikan kehitykseen vaikutti voimakkaasti 1970- ja 1980-lukujen yhteiskuntatieteiden niin kutsuttu kielellinen käänne eli sen tieteenfilosofisen käsityksen yleistyminen, että kieli rakentaa eikä vain heijasta todellisuutta ja on näin ollen tärkeä yhteiskunnallisen analyysin kohde. Tämä käsitys erosi merkittävästi aiemmista yhteiskuntatieteissä vallalla olleista positivistisista näkökulmista (ks. Johdannon luku 2 tk.). Käsitys vaikutti olennaisesti muun muassa näkemykseen siitä, mitä erilaisista kielellisistä aineistoista, kuten haastatteluista, voi tai kannattaa päätellä.

Metodisesti yhteiskuntatieteellisellä diskurssianalyysilla oli kuitenkin alkuvaiheissaan suhteellisen vähän yhteistä kielitieteellisen diskurssianalyysin kanssa. Käytännön työkaluja aineiston tulkintaan lainattiin sen sijaan esimerkiksi retoriikasta ja etnometodologisesta keskustelunanalyysista (ks. esim. Potter & Wetherell 1987). Osassa yhteiskuntatieteellistä diskurssianalyysia vaikutti – ja vaikuttaa edelleen – vahvana myös esimerkiksi foucault'laiseksi kutsuttu perinne (esim. Foucault 1970), jossa kielenkäyttöä tarkastellaan hyvin abstraktisti makrotasolta. Foucault'laisessa näkökulmassa diskurssi ymmärretään käsitteellisesti yhteiskunnallis-historiallisena ilmiönä, ja diskurssin analyysia voi edustaa esimerkiksi juuri jonkin tähdellisen yhteiskunnallisen ilmiön synnyn ja muotoutumisen analyysi.

## 2.3. Teoriataustan ja kielikäsityksen muovautuminen

Yhteiskuntatieteellisen diskurssianalyysin tärkeä anti kielitieteelliselle diskurssianalyysille koskee teoreettista taustaa. Diskurssianalyysin

teoriatausta – sen kielikäsitys – perustuu pitkälti yhteiskunnallisen ja sosiaalisen todellisuuden rakentumista käsittelevään tieteenfilosofiaan (esim. Berger & Luckmann 1966; Foucault 1970), jonka yhteiskuntatieteiden kielellinen käänne toi mukanaan. Tämä sosiaalisena konstruktionismina tunnettu teoria (tai teoriat) (ks. esim. Burr 2015 [1995]) on oppi todellisuuden sosiaalisesta rakentumisesta – ja siinä mielessä kielestä, että todellisuuden nähdään rakentuvan sosiaalisesti kielen avulla. Sosiaalinen konstruktionismi on sittemmin yleisesti hyväksytty diskurssianalyysin lähtökohdaksi myös kielitieteellisesti painottuneissa suuntauksissa.

Toinen tärkeä diskurssianalyyttista tutkimusta laajasti yhdistävä, suoremmin kieli-, kirjallisuus- ja viestintätieteiden omaan teoriakehitykseen liittyvä näkemys perustuu kielen ja merkityksen dialogisuuden ja kontekstuaalisuuden periaatteisiin, joita on muotoiltu esimerkiksi bahtinilaisessa dialogismissa (Bahtin 1981; 1986 [1953]) sekä sille monissa kohdin sukua olevassa jälkistrukturalismissa (esim. Barthes 1993). Kokoavasti sanoen dialogististen ja jälkistrukturalististen näkökulmien diskurssianalyysiin jättämä jälki koskee erityisesti sekä merkitysten dynaamisuuden, vuorovaikutuksellisuuden ja kontekstisidonnaisuuden korostusta että *tekijän* (tai *puhujan*) käsitteen haastamista siten, että kieltä ja kielenkäyttöä ei yksioikoisesti yhdistetä kieltä käyttävään yksilöön.

Diskurssianalyysin metodiikka on muutoinkin kuin teoriamielessä saanut tärkeitä vaikutteita ja tutkimuksellisia käsitteitä ja työkaluja kirjallisuudentutkimuksen piiristä. Esimerkiksi nimenomaan Barthes (1993) kuten myös Genette (1972) ovat antaneet diskurssianalyysille eväitä tulkita tekstin ja sen tekijyyden ja erilaisten kerronnan ”äänien” suhteita, ja ylipäätään narratiivien ja niiden rakenteiden tutkimus (ks. De Fina & Georgakapoulou 2015) on ollut ja on edelleen yksi keskeinen diskurssianalyytikkojen tutkimuskohde. Intertekstuaalisuus ja tekstien moniäänisyys (Kristeva 1986 [1966]; Bahtin 1981) ovat niin ikään merkittäviä analyysin kohteita ja analyyttisia kehikkoja diskurssianalyysille. Myös *genre* eli *tekstilaji* (ks. Mäntynen, Shore & Solin 2006; Heikkinen, Voutilainen, Lauerma, Tiililä & Lounela 2012) on keskeinen käsite kirjallisuudentutkimuksen lisäksi diskurssianalyysissa.

## 2.4. Diskurssianalyysin suomalaista historiaa

Diskurssianalyysi, nimenomaan nimillä *diskurssianalyysi* tai *diskurssintutkimus*, ei vakiintunut kovin nopeasti osaksi suomalaista kielentutkimusta. Erilaisia kielenkäytön ilmiöitä kyllä tarkasteltiin monilla kielentutkimuksen alueilla siten, että myös kansainvälisiä pragmaattisia ja diskurssianalyyttisia teorioita ja käsitteitä hyödynnettiin, mutta jälkeenpäin tarkasteltuna vaikuttaa siltä, että sen enempää pragmatiikka kuin diskurssianalyysikaan eivät hevin päässeet kielentutkimuksen alueen kaanoniin. Keskusteluntutkimus sen sijaan vakiinnutti oman asemansa esimerkiksi lukuisten suomen kielen tutkijoiden opinnäytetöissä jo 1980-luvun lopulta alkaen. 1990-luvulle ajoittuu joukko tekstintutkimuksen aluetta paaluttaneita suomen kielen tutkimuksia, jotka esittelivät ansiokkaasti diskurssintutkimuksellisia ajatuksia esimerkiksi kriittisestä diskurssianalyyttisesta tulkinnasta ja systeemis-funktionaalisen lingvistiikan soveltamisesta suomenkielisiin aineistoihin (esim. Karvonen 1995; Heikkinen 1999). Tutkimusalueen ja metodin nimitykset *diskurssianalyysi* tai *diskurssintutkimus* alkoivat yleistyä Suomessa kuitenkin vasta 2000-luvulle tultaessa (esim. Pälli 2003); joskin edelleen 2000-luvun alkupuolella näyttää kielentutkijoiden keskuudessa olleen yleinen tapa käytännössä määrittää tutkimuksen ala sen edustaman aineistotyypin kautta siten, että tekstiaineistoja sisältävät tutkimukset ovat olleet suomalaisittain tekstintutkimusta ja keskusteluaineistoja käyttävät tutkimukset keskusteluntutkimusta. Muuta puhuttua kieltä (esim. suulliset haastattelut) aineistoinaan käyttäneet tutkimukset perustuivat pääosin muihin, esimerkiksi sosiolingvistisiin metodeihin, mikä on sikäli yllättävää, että eri tieteenalat huomioon ottaen puhutun kielen haastatteluaineistot ovat hyvin tyypillinen diskurssianalyysin aineistomuoto.

Omaa kieltänsä diskurssianalyysin yleistymisestä metodisena otteena kertoo se, että ensimmäinen diskurssianalyyttinen suomenkielinen kielentutkimuksellisiin tarpeisiin kirjoitettu oppikirja, *Kurssi kohti diskurssia*, ilmestyi vasta 2009 (Pietikäinen & Mäntynen 2009). Tarpeesta suomenkieliselle oppikirjalle viestii, että kirjaan on vuoteen 2020 mennessä viitattu jo yli 1000 kertaa (Google Scholar).

Oppikirjatasolla kuitenkin ylivoimaisesti merkittävin yleinen diskurssianalyyttista tutkimusta suomeksi esittelevä teos lienee Arja Jokisen, Kirsi Juhilan ja Eero Suonisen *Diskurssianalyysin aakkoset* (Jokinen, Juhila & Suoninen 1993), jossa esitellään sosiaalisen konstruktionismin teoriaan perustuvaa diskurssianalyysia. Painokkaan sijan oppikirjassa sai erityisesti niin sanotun Loughborough'n koulukunnan (esim. Potter & Wetherell 1987) diskursiivisen (sosiaali)psykologian sen aikaisen työkalupakin sisältö. Samojen kirjoittajien *Diskurssianalyysi liikkeessä* (Jokinen, Juhila & Suoninen 1999) puolestaan heijasteli diskurssianalyysin kehityssuuntia edelleen Loughborough'n koulukunnan diskurssianalyytikoiden näkemyksiä seuraillen. Sama kirjoittajatiimi on jatkanut arvokasta, varsinkin yhteiskuntatieteilijöille suunnattua metodin esittelyä myös 2016 julkaistussa oppikirjassa *Diskurssianalyysi* (Jokinen, Juhila & Suoninen 2016).

Olennaista alan suomalaisesta kehityksestä on todeta, että 2000-luvulla diskurssintutkimukselliset otteet ovat tulleet Suomessa hyvin yleisiksi ihmisiä, kulttuuria ja yhteiskuntaa eri tavoin tarkastelevissa tieteissä. Kielitieteellinen perusta on monitieteiselläkin alueella edelleen erittäin tärkeä, ja esimerkiksi kielitieteisiin pohjautuvissa käsitteissä ja kielentutkimuksen piirissä syntyneissä teorioissa ja metodeissa on varmasti vieläkin paljon potentiaalia, jota yhteiskuntatieteellisesti suuntautunut diskurssianalyysi voi hyödyntää. Mainittakoon esimerkkinä syntaktis-semanttinen temaattisten roolien analyysi ja siihen liittyvä verbisemanttinen argumenttianalyysi, joka voisi soveltua uutena ja jännittävänäkin analyysikehikkona monenlaisiin toimintaa ja sen seurauksellisuutta pohtiviin analyyseihin. Vastaavasti esimerkiksi käyttäytymistieteissä, johtamis- ja organisaatiotutkimuksessa tai vaikkapa viestintä- ja mediatutkimuksessa on ollut jo pitkään vireillä tieteellisiä keskusteluja, jotka voivat edistää myös enemmän kielentutkimuksellisista lähtökohdista ponnistavaa diskurssianalyysia. Tällaisista voisi yhtenä esimerkkinä mainita niin sanotut käytäntöteoreettiset tutkimusotteet, joissa tutkimuksen kohteena olevaa ilmiötä tarkastellaan sosiaalisena käytäntönä ja tilanteisena sosiaalisena, kielellisenä, kehollisena ja materiaalisena toimintana.

Kyse on toki yleisemmästä kuin vain suomalaisesta kehityskulusta, ja toisaalta Suomessa kuten muuallakin on näkyvissä merkkejä siitä,

kuinka ilmiöpohjaisuus[I] osittain korvaa tieteenalaluokituksia esimerkiksi monitieteisten tutkimusryhmien muodossa. Diskurssin eli kielenkäytön ja sosiaalisen toiminnan yhteenkietoutuneisuuden analyysi on hedelmällinen maasto tällaiselle tieteenalarajat ylittävälle yhteistyölle.

**Mitä tutkitaan?**

Diskurssianalyyttisen tutkimuksen kirjo on niin suuri, että yksittäisiä tutkimuksia tai aiheita mainitsemalla on vaikea antaa kokonaiskuvaa siitä, mitä kaikkea diskurssianalyysilla voi tutkia. Laajalla otteella asiaa katsoen voitaisiin kuitenkin luetella muutamia suhteellisen tyypillisiä diskurssianalyysin kohdealueita:

- yhteiskunnalliset epätasa-arvoiset käytännöt ja syrjintä
- poliittinen retoriikka
- yhteiskuntaelämään liittyvä kielenkäyttö ja toiminta (esim. kunnallishallinnon ja sen eri alainstituutioiden kielenkäyttö)
- yritysten sosiaalisiin käytäntöihin liittyvä retoriikka
- instituutioiden kertomukset tai muut selonteot itsestään (diskurssianalyysia voi hyvin tehdä vaikkapa yritysten tilinpäätöksistä!)
- inhimillisten toimijoiden kertomukset tai muut selonteot tutkijaa kiinnostavasta yhteiskunnallisesta tai sosiaalisesta ilmiöstä
- kielen tai sen käytön rooli osana jotain tiettyä joko institutionaalista (esim. muodolliset ammatilliset kokouskeskustelut) tai jotain muuta sosiaalista aktiviteettia (esim. kaveriporukan keskustelut kotibileissä).

Listan jatko on diskurssianalyyttisesta mielikuvituksesta kiinni. Mielikuvitus on kuitenkin usein sidoksissa havaintoihin: hyvä diskurssintutkija huomaa säännönmukaisuuksia kohtaamassaan kielenkäytössä siten, että hän samaan aikaan näkee sosiaalisia toistuvuuksia ja hypoteesimaisesti peräti selitysmalleja kohtaamaansa kielenkäyttöön liittyen.

Annamme kaksi lyhyttä esimerkkiä. Eräs tuttavamme totesi kerran mielipiteenään kokouskäyttäytymisestä, että ihmiset puhuvat ja käyttävät mielipiteitä organisatorisissa kokouksissa pääosin sen vuoksi, että haluavat ilmoittaa, että ovat olemassa ja ovat tärkeitä. Tuttavamme ilmaisi

asian analogialla, jossa puheenvuoro kokouksessa voisi koostua hänen mielestään esimerkiksi seuraavasti: ”Minä nyt puhun, kuulkaa, minä puhun, se olen minä joka puhun ja minulla on tässä asiaa ja nimenomaan minä tässä nyt puhun”. Tällainen havainto – jonka sivumennen sanoen oli tehnyt suhteellisen paljon erilaisissa kokouksissa istunut henkilö – voisi olla johdettavissa kiinnostavasti systemaattiseen diskurssianalyyttiseen tutkimukseen kokouspuheenvuoroista. Toinen tuttavamme, myös hän vaativassa asiantuntija- ja esimiestehtävässä, kertoi puolestaan kerran työyhteisöllisenä havaintonaan, kuinka työyhteisöstä löytyy puhetapa ja ihmistyyppi, jolle on leimallista negatiivisuuden etualaistaminen, mikä näkyy sellaisessa kielenkäytössä kuin *tääl ei oo ennenkään*, *kukaan ei koskaan*, *kaikki aina*, *ennen ei* tai argumentaation suunnasta riippuen *ennen aina*. Havainto olisi varsin maukas työyhteisöpuheen analyysin kannalta.

# 3. Diskurssianalyysin menetelmiä

Diskurssianalyysin monenlaisten suuntausten lähtökohdat ja tavoitteet näkyvät erilaisissa aineisto- ja menetelmävalintoissa. Seuraavassa esittelemme näitä lyhyesti.

## 3.1. Aineistot, suuntaukset ja lähtökohdat

Koska periaatteessa kaikenlaiset kielenkäytön ja viestinnän muodot ovat mahdollisia diskurssianalyysin kohteita, käytettävän aineiston määrittävät pääosin tutkimuksen tavoitteet ja tutkimuskysymykset. Osin aineiston valinta riippuu myös kyseessä olevan diskurssianalyysin suuntauksen perinteistä – toiset suosivat kirjallista aineistoa, toiset keskustelua. Aineistoa voi myös täydentää haastatteluilla tai etnografisella havainnoinnilla, jotta saavutetaan parempi ymmärrys kontekstista. Nykyään

pidetään enenevässä määrin myös tärkeänä, että analyysissa ei keskitytä pelkästään kielen analyysiin vaan huomioidaan multimodaalisuus. Erityisesti internetin ja sosiaalisen median jatkuvasti kasvavan merkityksen myötä tämän näkökulman olennaisuus korostuu: suuri osa päivittäin kohtaamistamme teksteistä sisältää myös kuvallisia elementtejä ja videota. Tässä yhteydessä on syytä tuoda esiin sekin, että vaikka diskurssianalyysi on aina joltain osin laadullista, nykyisin on suhteellisen yleistä yhdistää analyysimenetelmään myös määrällisiä korpustyökaluja (korpusten käytöstä ks. Laadullinen aineistopohjainen kielentutkimus; Korpusaineistot; Määrällinen korpuslingvistiikka; Historiallinen korpuslingvistiikka tk.), tietenkin niin, että ne palvelevat diskurssianalyysin tavoitteita (ks. esim. Baker ja muut 2008; Lillqvist 2019). Osuva nimitys korpuksia hyödyntävälle diskurssintutkimukselle onkin *korpusavusteinen diskurssianalyysi*.

Suuremmat aineistot ja korpusmenetelmien tarkoituksenmukainen soveltaminen voivat eittämättä auttaa säännönmukaisuuksien löytämisessä ja niiden osoittamisessa (ks. esim. Mautner 2016) – mikäli siis tutkimuskysymys on asetettu siten, että tarkasteltavasta ilmiöstä on tarkoitus osoittaa säännönmukaisuutta tai yleisyyttä. Tyypillisesti diskurssianalyyttista otetta aineiston tulkintaan luonnehtii kuitenkin joko induktiivinen, yksittäisistä havainnoista kohti yleisempiä tulkintoja kohdistuva päättely tai abduktiivinen, teoriasidonnainen päättely, jossa vähälukuinenkin joukko havaintoja voi olla kiinnostava, kun se liittyy esimerkiksi olemassa olevan teoreettisen keskustelun käsitteisiin ja johtoajatuksiin.

Diskurssintutkimuksen erilaisia suuntauksia voidaan erotella karkeasti sen mukaan, millaisista ilmiöistä ne ensi sijassa haluavat tuottaa tietoa. Toiset pitävät tärkeimpinä sosiaalisia tai yhteiskunnallisia ilmiöitä, joita erilaisten kielenkäytön tapojen analysointi voi valottaa, toiset puolestaan ovat erityisen kiinnostuneita kielenkäytön ilmiöistä aidoissa sosiaalisissa käyttökonteksteissa.

Esimerkiksi kriittisessä diskurssianalyysissa keskitytään yhteiskunnallisesti merkittäviin ilmiöihin, kuten ideologioihin ja valtaan, sekä siihen, kuinka nämä synnyttävät ja ylläpitävät epätasa-arvoa. Tämän vuoksi tutkimusprosessissa lähdetään usein liikkeelle tietystä

yhteiskunnallisesta ongelmasta ja kerätään tätä mahdollisimman hyvin havainnollistavaa aineistoa. Vaikka kriittistä diskurssianalyysia luonnehtii melko yksityiskohtainen kielellinen analyysi (Fairclough 2003) ja vaikka sen kautta toki saadaan tietoa myös kielen merkityspotentiaalista ja kielen rakenteista käytössä, pääasialliset tavoitteet ja tutkimuskysymykset liittyvät siis yhteiskunnallisiin ja sosiaalisiin ilmiöihin. Kriittisessä diskurssianalyysissa voidaan esimerkiksi tutkia, miten maahanmuuttajista puhutaan ja kirjoitetaan, ja valita aineistoksi mediatekstejä tai poliittisia puheita, joissa maahanmuuttoa käsitellään. Toisaalta aineistonkeruun lähtökohdaksi voidaan ottaa muukin kiinnostavaksi ja merkittäväksi koettu asia tai henkilö, esimerkiksi Donald Trump ja hänen retoriikkansa tai EU:n toimielimissä käytetty argumentaatio. Kriittisen diskurssianalyysin pyrkimys yhteiskunnallisesti merkittävien kysymysten äärelle kuitenkin tarkoittaa sitä, että myös tarkasti rajatut analyysin kohteet kertovat jostain laajemmasta ilmiöstä (kuten populismista, rasismista tai uusliberalismista).

Toisaalta taas esimerkiksi diskursiivisen psykologian mukainen tai siitä ponnistava diskurssianalyysi keskittyy usein sosiaalipsykologisiin ilmiöihin, kuten siihen, kuinka psykologisia tiloja ja tunteita käytetään ja tuotetaan teksteissä tai puheessa. Tutkija voi kohdistaa analyysin julkisiin, institutionaalisiin (esim. työpaikoilla tapahtuviin) tai arkipäiväisiin teksteihin tai keskusteluihin, joissa esimerkiksi psykologisoivia kuvauksia ja termejä tai yleisemmin vain ”psykologisointia” käytetään. Tavoitteena ja tutkimuksen motivaationa on ennen kaikkea ymmärtää psykologisten kuvausten ja siihen liittyvän retoriikan sosiaalista seurauksellisuutta (Wiggins & Potter 2008). Toisaalta tavoite voi olla myös valaista kielen rakenteisiin kiteytynyttä psykologisuutta.

Kielenkäytön ilmiöihin keskittyvät usein esimerkiksi tekstilajeja tai retoriikkaa tutkivat diskurssianalyysit. Tutkija voi silloin olla kiinnostunut ensi sijassa siitä, millaiset piirteet tiettyjä tekstejä tai tietynlaisia tekstejä luonnehtivat, ja kohdistaa analyysin kärjen näiden piirteiden tarkasteluun. Silti näissäkin lähestymistavoissa voi kyllä paikantaa itselleen tutkittavaksi yhteiskunnallis-sosiaalisen ilmiön, kuten sen, mitä tietyillä teksteillä tehdään ja miten ne ohjaavat toimintaa ammatillisissa tilanteissa (esim. Devitt 1991; Bazerman 1997). Sama pätee

intertekstuaalisuuteen ja narratiivisuuteen, jotka molemmat varsin usein esiintyvät diskurssianalyysin metodeina joko itsenäisesti tai muihin analyysitapoihin kytkeytyneinä. Intertekstuaalisuuden tutkija voi painottaa joko funktiota eli sitä, mitä intertekstuaalisuus saa aikaan, tai rakennetta eli sitä, millaisin keinoin intertekstuaalisuus esimerkiksi tietyntyyppisissä teksteissä rakentuu. Narratiiveja eli kertomuksia tutkiva voi niin ikään olla enemmän kiinnostunut joko kertomusten rakenteellisista ominaisuuksista tai niiden funktioista. Vastaavaa voi sanoa esimerkiksi multimodaalisista ja vuorovaikutuslingvistisistä analyysitavoista, jotka nekin voivat keskittyä enemmän joko rakenteeseen tai käyttöön ja käytön sosiaalisiin vaikutuksiin.

Olennaista on siis, että kyse on painotuksista. Hedelmällisempää kuin esitellä sellaisenaan erilaisia suuntauksia ja koulukuntia onkin hahmottaa diskurssianalyysin kenttää erilaisten tavoitteiden ja työkalujen yhdistelmien kautta.

## 3.2. Analyysin työkaluja

Analyysimenetelmän valintaa ohjaavat tutkimuksen tavoitteet, aineistotyyppi ja aineiston määrä. Analyysissa voidaan keskittyä yhteen hyvinkin tarkkarajaiseen näkökulmaan tai vain yhteen analyyttiseen käsitteeseen, kuten narratiivisuuteen tai metaforiin. Usein analyysin tekijä kuitenkin käyttää useampaakin työkalua, jotka voivat liittyä erilaisiin kielen ja vuorovaikutuksen keinoihin, tekstuaaliseen ja sosiaaliseen kontekstiin tai esimerkiksi siihen, millaisista osista teksti tai vuorovaikutustilanne koostuu, mistä aiheista puhutaan tai missä suhteessa asiat saavat huomiota.

Olennaista on, että menetelmän valintaan vaikuttavat analyysin tekijän havainnot siitä, mitä aineistossa on. Joskus sellainenkin tapa voi toimia, että tutkija valitsee etukäteen jonkin analyyttisen työkalun tai kehikon ja yksinkertaisesti vain järjestelmällisesti tutkii oman aineistonsa tuon työkalun kautta. Tavanomaisempaa ja yleensä toimivampaa on kuitenkin, että tutkijalla on mielessään alustava tutkimusongelma ja tutkimusasetelma, joita vasten hän tekee ensin alustavia havaintoja

siitä, millaiset kielenkäytön ilmiöt korostuvat juuri hänen aineistossaan. Todennäköisesti nuo havainnot liittyvät johonkin tai joihinkin käsitteistä ja ilmiöistä, joita esittelemme taulukossa 1 ja luvuissa 3.3. ja 3.4. käsittelemissämme kahdessa esimerkissä. Niinpä näitä analyyttisia käsitteitä ja niihin liittyvää teoreettista tietoa sekä esimerkkejä muiden tutkimuksista käytetään oman analyysin perustelemiseen ja systematisoimiseen, ja samalla oma analyysi osallistuu diskurssianalyyttiseen keskusteluun kyseisestä käsitteestä tai työkalusta.

Taulukkoon 1 on koottu muutamia diskurssianalyysissa usein käytettyjä analyyttisia käsitteitä selityksineen. Taulukossa kerromme käsitteiden tutkimusalaperustan, annamme selityksen, jossa kuvaamme käsitettä lyhyesti nimenomaan sen diskurssianalyyttisen hyödynnettävyyden kannalta, sekä lisäksi koetamme muutamalla esimerkinomaisella mahdollisella tutkimuskysymyksellä valottaa sitä, miten käsite voisi liittyä erilaisiin tutkimuskysymyksiin. Seuraavaksi annamme kaksi erilaista esimerkkiä aineistotyypeistä ja mahdollisista analyysin näkökulmista niihin.

## 3.3. Esimerkki kirjalliseen aineistoon perustuvasta diskurssianalyyttisesta tulkinnasta

Ensimmäinen esimerkki, jolla havainnollistamme diskurssianalyyttisia menetelmiä käytännössä, on eduskunnan tiedote kyselytunnista 21.9.2017 (ks. esimerkki 1). Teksti keskittyy pienituloisten eläkeläisten asemaa koskevaan keskusteluun eduskunnassa, ja se voisi siten olla osana aineistoa esimerkiksi tutkimuksessa, jonka tavoitteena olisi selvittää, miten köyhyydestä puhutaan politiikassa tai laajemminkin yhteiskunnassa. Tällöin aineisto todennäköisesti sisältäisi eduskunnan tiedotteiden lisäksi paljon muunkinlaisia tekstejä.

Toisaalta tekstiä voisi tarkastella siitä näkökulmasta, millainen tekstilajinsa edustaja se on: silloin muu aineisto koostuisi myös tiedotteista. Näin voitaisiin vertailla tätä tekstiä muihin tiedotteisiin, joko esimerkiksi tiedotteisiin ylipäätään tai nimenomaan muihin eduskunnan kyselytunteja koskeviin tiedotteisiin. Vertailemalla tekstiä muihin vastaaviin

Taulukko 1. Diskurssianalyysin käsitteitä.

| Käsite | Perusta | Selitys | Analyysin kysymyksiä |
|---|---|---|---|
| Fokalisointi (fokus, fokalisaatio) | Teksti-lingvistiikka, kirjallisuuden-tutkimus, narratologia | Näkökulma ja kerronnan perspektiivi (erityisesti Genette 1972). | Kenen näkökulmasta kuvaus on tehty? Millä keinoilla näkökulmia kielellisessä kuvauksessa luodaan ja vaihdellaan? Mitä niillä tehdään? Mitä valittu näkökulma tuo esiin ja korostaa? |
| Teema/reema, teema-progressio, informaatio-rakenne | Teksti-lingvistiikka | Informaation esittäminen tekstissä vanhana tai uutena ja juonen ja argumentaation eteneminen temaattisesti (Daneš 1974). | Miten asioita esitetään uusina tai jo tunnettuina ja mitä tämä merkitsee kuvaukselle (esim. sen näkökulmalle)? Millainen teeman eteneminen luonnehtii esimerkiksi tietynlaisia tekstejä? |
| Makro-propositio | Teksti-lingvistiikka, kognitiivinen diskurssi-analyysi | Semanttisten makrotason sääntöjen (esim. yleistäminen, poistaminen, yhdistäminen) kautta luodut tiivistykset. Kukin tekstikappale on kuvattavissa makro-propositiolla, joka tiivistää kappaleen esim. yhdeksi lauseeksi. (Esim. van Dijk 1977.) | Millaiset makropropositiot luonnehtivat tarkasteltavia tekstejä? Mitkä ovat keskeiset aihepiirit? Miten tarkasteltavien tekstien logiikka tai argumentaatio rakentuu makropropositiosta toiseen? |
| Koheesio ja koherenssi | Teksti-lingvistiikka | Kielellisen esityksen sidosteisuus, jota tuotetaan esimerkiksi kielellisillä sidoskeinoilla (esim. yhdistävät konnektiivit, toisto tai teema–reema-vaihtelu). Koherenssi ymmärretään yleensä laajemmin lukijan tekstiin heijastamana tulkinnallisena logiikkana, jota vasten hän tulkitsee tekstin mielekkääksi. Koherenssi koostuu erilaisista kontekstuaalisista, sosiokulttuurisista ja kielellisistä yhteyksistä. (Halliday & Hasan 1976.) | Miten sidosteisuuden keinot tietyssä kontekstissa käytettynä rakentavat tekstin tulkinnallista logiikkaa? Millä tavoin koheesiokeinot ohjaavat kulttuurisen tulkinnan tekemistä esimerkiksi asioiden syy–seuraus-suhteista tai yhdistymisestä toisiinsa? |
| Prosessi-tyypit | Teksti-lingvistiikka, systeemis-funktio-naalinen kielitiede | Kielellistä esitystä voidaan katsoa lause-, virke-, tai myös laajemmalta tasolta eritellen siitä kielellisiä prosessityyppejä ja niihin kytkeytyvien osallistujien ja olosuhteiden luonnetta. Pääprosesseja ovat materiaaliset (tekeminen), mentaaliset (kognitiivinen toimintaa), behavioraaliset (käyttäytyminen), verbaaliset (sanominen), relationaaliset (oleminen ja omistaminen), eksistentiaaliset (olemassaolo tai olevaksi tulo) prosessit. (Halliday 1994.) | Millaiset prosessit luonnehtivat tietynlaisia tekstejä – tai tietynlaisista aihepiireistä kirjoittamista tai puhumista? Millaisia eroja on esimerkiksi tiettyjen henkilöiden tai ryhmien käyttämissä (tai heistä käytetyissä) prosessikuvauksissa? |

| Käsite | Perusta | Selitys | Analyysin kysymyksiä |
|---|---|---|---|
| Temaattiset (semantti-set) roolit | Syntaksi, semantiikka | Sijakielioppiin (Fillmore 1968) perustuva näkemys kielellisen (lausetason) kuvauksen funktionaalista merkitystä kantavista yksiköistä, joilla on "sija" suhteessa verbin ilmaisemaan toimintaan. Rooleja erotellaan hieman eri määriä ja eri tavoin, mutta tyypillisiä ovat esimerkiksi agentti eli tekijä, instrumentti eli väline, patientti eli vastaanottaja, kokija (experiencer), hyötyjä eli benefaktiivi, lähde (source) ja määränpää (goal). (Fillmore 1968; Frawley 1992.) | Miten erilaiset sijaroolit toimivat merkitys-resursseina kielellisessä kuvauksessa? Mitä rooleja kuvaus hyödyntää? Millaista systematiikkaa ja millaisia sosiaalisia tulkintoja kielellisen kuvauksen rooleihin liittyy? |
| Leksikaaliset merkitys-suhteet | Seman-tiikka, retoriikka | Sanatason merkityssuhteet, eli jostain aiheesta puhuttaessa käytössä olevat sanastolliset valinnat ja niiden semantiikka. Diskurssianalyytikko voi katsoa merkityssuhteita (esim. polysemiaa, synonymiaa tai antonymiaa) valintoina, jotka nostavat esille erilaisia merkityksiä ja ohjaavat tulkintaa. Merkityssuhteisiin kuuluvat myös assosiatiiviset ("sivu")merkitykset, joita tiettyyn valintaan liittyy (esim. konnotaatiot, kollokaatiot tai affektisuus) (Leech 1974). | Millaisen näkökulmia ja merkityksiä – diskursseja – tekstin leksikaalisiin valintoihin kytkeytyvät merkityssuhteet tuovat esiin? Voidaanko esimerkiksi eri toimijoiden tekstien välillä huomata sellaisia diskurssin eroja, jotka näkyisivät sanastollisissa merkityssuhteissa? |
| Modaalisuus | Semantiikka | Kielen keinot, jotka ilmentävät puhujan asenteita puhuttavaan asiaan: esim. puhua asioista mahdollisina, toivottavina tai epätoivottavina, varmoina tai epävarmoina, pakollisina, luvallisina tai välttämättöminä. (Nuyts 2005.) | Millaisia esimerkiksi tietämiseen (episteeminen mod.), tahtomiseen (deonttinen mod.) tai kykyyn (dynaaminen mod.) liittyviä keinoja käytetään vuorovaikutussuhteiden rakentamiseen? |
| Varaukset (hedges) ja dialogi-partikkelit (diskurssi-partikkelit) | Pragma-tiikka | Tiedon tai toiminnon merkitseminen epävarmaksi tai ilmaisun sisällön modifiointi, esimerkiksi lieventäminen. Lähellä varauksia ovat myös dialogi- tai diskurssipartikkeleiksi, pragmaattisiksi partikkeleiksi tai diskurssin merkitsimiksi nimitetyt pikkusanat (*no, ai, tota*), jotka tilanteisesti luonnehtivat puhujan asenteita tai joilla puhuja osoittaa tulkintaansa aiemmin puheena olleesta. (Maschler & Schiffrin 2015.) | Millaisten toimintojen kuvauksiin liitetään varauksia? Millaisia sosiaalisia selityksiä varauksille voidaan antaa? Miten ne muuttavat merkityksiä tilanteessa? |

| Käsite | Perusta | Selitys | Analyysin kysymyksiä |
|---|---|---|---|
| Subjekti-asema | Psyko-analyysi, psykologia, narra-tiivinen tutkimus | Subjektius viittaa sekä näkökulmaisuuteen asioiden esittämisessä että rooleihin, joita eri ihmisryhmille yhteiskunnassa tarjotaan tai joita puhujat itse tämän seurauksena tiedostamattomastikin omaksuvat. Subjektius on yhteydessä subjektiviteettiin: omaksuttuaan tietynlaisen position henkilö näkee, kokee ja kertoo tämän position kautta hyödyntäen siihen kuuluvia metaforia, sanastoa, henkilötoimijuuksia, selontekomalleja ja skriptejä sekä tarinalinjoja. (Davies & Harré 1990.) | Millaisia subjektiviteetteja tai subjektipositioita on luettavissa esimerkiksi tietyn ihmisen tiettyä asiaa koskevasta kuvauksesta, ja miten erilaiset kielenkäytön keinot tukevat tulkintoja tietynlaisesta positiosta? |
| Diskurssi-identiteetti | Yhteiskunta-tieteellinen diskurssin-tutkimus, psykologia, vuorovaiku-tuslingvis-tiikka | Yleensä diskurssianalyysissa identiteettiin viitataan diskurssi-identiteettinä eli sellaisina puhujan valintoina, joiden kautta puhuja(t) rakentavat suhdettaan puhuttavaan asiaan, tai rakentavat itseään, kuulijaa tai ylipäänsä toisia ihmisiä tietynlaisina ihmisinä. (Benwell & Stokoe 2006.) | Millaisiksi itse, puhekumppani(t) tai muut ihmiset kuvataan ja miten tämä vaikuttaa osallistumisen tapaan tai mahdollisuuksiin esimerkiksi tilanteisessa vuorovaikutuksessa? Millaisia säännönmukaisuuksia on esimerkiksi tietyn ammattiryhmän edustajien puheessa itsestään? Millaisia säännönmukaisuuksia on puhetavoissa, jotka koskevat esimerkiksi jotain yhteiskunnallista ryhmää? |
| Diskursiivinen legitimaatio ja luonnollistaminen | Kriittinen diskurssi-analyysi | Tavat tehdä asioita oikeaksi, järjelliseksi ja hyväksyttäväksi kielen ja argumentaation keinoin. Esimerkiksi mielipide voidaan oikeuttaa ja tehdä luonnolliseksi niin, että se esitetään virallisena ja siten legitiiminä (autorisointi), järkevänä (rationalisointi), moraalisesti oikeana (moraalisesti oikein ajattelevien pakollisesti hyväksymä) tai kytkien tarkasteltavan asian tai ilmiön menneisyyteen tai tulevaisuuteen tarinallisesti (mythopoiesis) (van Leeuwen 2007). | Millaisia diskursiivisen legitimaation tapoja tietyt tekstit hyödyntävät? Millaisia legitimaatiotapoja tietyt tekstit hyödyntävät ja millaista todellisuutta ne tuottavat? |
| Etualaistaminen/taka-alaistaminen | Kriittinen diskurssi-analyysi/semiotiikka | Asioiden esittäminen ja näkökulman valinta joko niin että ne ovat ”etualalla”, kiinnittävät huomiota, tai jäävät vastaavasti taustalle – tätä voidaan tarkastella niin kuvista, liikkuvasta kuvasta kuin teksteistäkin. (Fairclough 2001 [1989].) | Mihin huomio kiinnittyy ja millä keinoin? Mitkä asiat rakentuvat tekstin tai kuvan pääasioiksi tai vastaavasti vähemmän olennaisiksi? Kenen näkökulma kuvauksessa korostuu? |

| Käsite | Perusta | Selitys | Analyysin kysymyksiä |
|---|---|---|---|
| Selonteot (accounts) | (Uusi) retoriikka, retorinen psykologia | Tapa tehdä selkoa asioista ja puolustautua esimerkiksi kiistämällä (denial), oikeuttamalla (justification), selittämällä (explanation), esittämällä (teko)syitä (excuses) tai pyytämällä anteeksi (apologies) (Potter & Wetherell 1987). | Millainen selontekotapa luonnehtii analysoitavaa puhetta/tekstejä? Millaisia selontekotyyppejä eri toimijat käyttävät? Miten tarkemmin ottaen jokin selontekotyyppi rakentuu tietyssä kontekstissa? |
| Retoriset rakenteet ja argumentaatio | Klassinen retoriikka, uusi retoriikka, retorinen (sosiaali) psykologia | Argumentaatiota voidaan tarkastella käsitteiden ethos, pathos ja logos (puhujan uskottavuuteen, tunteeseen ja järkeen vetoaminen) kautta. Klassisesta retoriikasta asti on kiinnitetty huomiota myös tarkempiin argumentoinnin keinoihin, joista diskurssianalyysissa hyödynnettyjä ovat esimerkiksi seuraavat ilmiöt/käsitteet: prolepsis (disclaiming; vastatodistelu) eli puhujan suuntautuminen niihin argumentteihin, joita hänen oma puheensa potentiaalisesti herättää (en ole rasisti, mutta...); dispositionaalinen-vastadispositionaalinen puhe eli puhujan tapa osoittaa yhtyvänsä kulttuuriseen jäsennykseen tai esimerkiksi eroavansa siitä esittäen toisen disposition (esim. kun tulee ikää, on luontevaa ajatella eläkettä vs. kun tulee ikää, on luontevaa ajatella eläkettä, mutta minä en...); kvantifikaatioretoriikka eli lukumääriin kuten prosentit tai muutoin määrällisiin luonnehdintoihin tukeutuminen (esim. kaikki, ei kukaan); skriptit (script formulations) eli kielen tiettyjen rakenteiden (kuten jos ...niin) avulla luodut loogiset päätelmät ja hypoteesit (esim. Edwards & Potter 1992; Potter 1996). | Millaista retoriikkaa analysoitavassa tekstissä käytetään? Miten asiaintiloja tekstissä tai puheessa tuotetaan todeksi tai perustelluiksi tai oikeiksi retoristen valintojen kautta? Miten jokin tietty retorinen rakenne esiintyy tietyillä yhteiskunnallisilla alueilla tai tietyissä sosiaalisissa tilanteissa? |
| Implisiittiset oletukset | Kriittinen diskurssianalyysi, pragmatiikka | Kielellisen kuvauksen taustalla luonnollisina "näin asiasta ajatellaan" -tyyppisinä ajatusmalleina olevat oletukset siitä, mitä on olemassa (eksistentiaaliset oletukset), mikä on, voi olla tai tulee olemaan asiaintila (propositionaaliset oletukset) ja mikä on hyvää tai haluttavaa (arvo-oletukset). On myös loogisia implikaatioita, jotka voidaan päätellä suoraan ilmaisusta, ja konversationaalisia implikatuureja, jotka voidaan päätellä keskustelun yleisistä periaatteista eli maksiimeista. (Grice 1975; Fairclough 2003.) | Millaisia taustaoletuksia analysoitavaan tekstiin liittyy? Mitkä asiat siinä nähdään itsestään selvinä? Millaisia muutoksia tai eroja tekstien taustaoletuksissa on historiallisesti tai paikallisesti? |

| Käsite | Perusta | Selitys | Analyysin kysymyksiä |
|---|---|---|---|
| Kategorisaatio ja partikularisaatio | Retorinen (sosiaali) psykologia, kognitiivinen psykologia | Asioiden, ilmiöiden tai ihmisten (ihmisten = sosiaalinen kategorisaatio) niputtaminen yhteen esimerkiksi retorisia tarkoituksia varten, tai vastaavasti kielen keinojen käyttö asioiden, ilmiöiden tai ihmisten ainutkertaistamiseen (partikularisoimiseen) (Pälli 2003). | Millaiset kategorisoinnin tavat luonnehtivat esim. tietyistä yksilöistä tai ryhmistä käytettyjä puhetapoja? Miten tietynlaisissa kielenkäyttötilanteissa tai tiettyjen kielenkäyttäjäryhmien toimesta kategorisoidaan? |
| Puheaktityypit | Pragmatiikka | Kielen käyttöön kiteytyneet tulkinnat tehtyjen puhetekojen merkityksistä – esimerkiksi että komissiivilla (sitoutumista osoittavalla ilmauksella) puhuja sitoutuu (tulen tänään teille) tai deklaratiivilla (todellisuutta muuttavalla ilmoituksella) puhuja muuttaa olosuhteita (julistan väitöstilaisuuden alkaneeksi). Käytännössä puheaktit ovat usein mutkikkaita ja epäsuoria: jos puhujalla olisi esimerkiksi aie/halu saada kuulija avaamaan ikkunan, hän voi käyttää käskyä tai kysymystä, joiden tilanteinen tulkinta puolestaan voi perustua esimerkiksi puhujien väliseen hierarkiaan. (Austin 1962; Searle 1975.) | Millaisia puhetekoja tietyissä tilanteissa tehdään ja miten ne vaihtelevat esimerkiksi ajassa, paikassa, kulttuureittain tai ihmisittäin? Millaiset puheteot luonnehtivat tiettyjä toimintoja tai esimerkiksi institutionaalisia vuorovaikutustilanteita? |
| Kasvojen uhkaaminen ja uhkaavuuden lieventäminen | Pragmatiikka | Tietyt puheteot voidaan nähdä kuulijan kasvoja uhkaavina tai muutoin sosiaalisesti ongelmallisina tilanteessa. Tämän vuoksi on kielenkäyttöön kiteytyneitä keinoja ilmaista periaatteessa uhkaava tai ongelmallinen puheteko kasvoja uhkaamatta tai ainakin uhkaa lieventäen, käyttää esimerkiksi kysymystä käskyn sijasta. (Brown & Levinson 1978.) | Miten tietynlaisissa konteksteissa kasvoja uhataan tai säilytetään ja millä vaikutuksilla? Kuinka esimerkiksi erilaiset puhujaryhmät lieventävät kasvojen uhkaa – ja mihin tarkoituksiin? |
| Persoona | Kielioppi, pragmatiikka | Diskurssianalyyttinen mielenkiinto persoonaan liittyy ennen kaikkea siihen, mitä persoonaviittauksilla teksteissä tai puheessa tehdään ja mitä ovat ne mahdolliset sosiaaliset syyt, joilla tiettyjä persoonia käytetään tai kaihdetaan. Yksi diskurssintutkijoiden kiinnostuksen kohteista on ollut passiivi, johon liittyen on keskusteltu esimerkiksi persoonan (usein tietoisestakin) hämärtämisestä ja toisaalta persoonaisuuden monimuotoisuudesta. (Pälli 2003.) | Keitä sisältyy esimerkiksi monikollisten pronominiviittausten alaan? Millaisessa kontekstissa puhuja puhuu minä-muodossa? Millaisia merkityksiä nostetaan esiin esimerkiksi geneerisellä persoonalla (suomessa nollapersoona tai puhekielinen *sä*) tai passiivilla? |

| Käsite | Perusta | Selitys | Analyysin kysymyksiä |
|---|---|---|---|
| Deiksis | Kielioppi, pragmatiikka | Kielen deiktisten tai demonstratiivisten viittaaminen joko kielellisen esityksen ulkopuolella (eksofora) tai toiseen kohtaan kielellisessä esityksessä (endofora). Diskurssianalyytikolle deiktiset viittaukset ovat kiintoisia tilanteisessa käytössään, jossa ne voivat tuoda esiin erilaisia näkökulmia esimerkiksi puhetilanteen henkilöiden suhteisiin (sosiaalinen deiksis ja persoonadeiksis) tai kielellisesti kuvatun toiminnan tapahtuma-aikaan (tempusdeiksis). (Levinson 2004.) | Millaisia säännönmukaisuuksia ja niihin liittyviä sosiaalisia selityksiä ilmausten deiktisiin ja demonstratiivisiin ilmauksiin liittyy? Miten tietyt tekstit tai muut kielelliset esitykset hyödyntävät deiksistä? |
| Tekstilaji (genre) | Kirjallisuudentutkimus, tekstien ja ammatillisen viestinnän tutkimus | Sosiaalinen toimintatyyppi, sosiaalisesti vakiintunut tapa tehdä asioita tietyllä viestinnällisellä tavalla. Esimerkiksi tiedote tiedottaa; kirjallisella tuotteella ”tiedote” on siis vakiintunut viestinnällinen päämäärä ja sen yleinen ymmärrys, sekä niihin liittyvästi vakiintuneet tyypilliset kielelliset valinnat. (Mäntynen, Shore & Solin 2006.) | Miten tekstilajeja käytetään erilaisiin tavoitteisiin? Miten asioita tehdään käyttäen tietoa tekstilajeista? Miten tekstilajit liittyvät ammatilliseen osaamiseen ja ammatin sosiaalisiin käytäntöihin? |
| Intertekstuaalisuus | Kirjallisuudentutkimus, kriittinen diskurssianalyysi | Kielellisen esityksen tavat hyödyntää muita esityksiä, luoda niihin yhteyksiä ja muodostaa yhteyden kautta merkityksiä. Esimerkiksi lainaaminen tai suora viittaaminen, kieliopillinen suora tai epäsuora esitys. Lähitermi interdiskursiivisuus viittaa diskurssien tapaan hyödyntää tai kaiuttaa muita diskursseja, puhetapoja. Liittyy läheisesti moniäänisyyden käsitteeseen (polyfonia), eli useiden erilaisten näkökulmien tai ”äänten” esiintymiseen samassa tekstissä (Bahtin 1981; Fairclough 1992). | Miten esimerkiksi tietynlaisissa teksteissä taustalla olevat ajatusmallit siirtyvät toisenlaiseen ympäristöön, toisenlaisiin teksteihin intertekstuaalisuuden keinoin? Miten intertekstuaalisuus tuo tekstiin tai muuhun kielenkäyttöön moniäänisyyden ja mitä seurauksia moniäänisyydellä on? |
| Narratiivit ja esimerkit | Kirjallisuudentutkimus, yhteiskuntatieteet, retoriikka | Tarina tai kertomus ja niiden rakenne ja funktiot (mitä saavat aikaan kielellisen esityksen tulkinnan kannalta). Tarkastelussa voidaan keskittyä kertomuksen rakenneyksiköihin rakenteeseen tai tarkkailla kertomuksellisuutta sen funktioiden kautta. Kertomuksina tai narratiiveina voidaan pitää myös laajemman kielellisen interaktion – esimerkiksi keskustelun – sisällä olevia ”minikertomuksia” (microstories), joilla voidaan tulkita olevan omia funktioitaan osana sitä laajempaa toimintoa, jossa ne esiintyvät. (De Fina & Georgakapoulou 2015.) | Millaisia kertomuksia tietynlaiset kielelliset esitykset hyödyntävät? Miten kertomukset vastaanotetaan keskustelutilanteessa tai muussa kielenkäyttötilanteessa? Mitä kertomusrakenteita ja millä vaikutuksilla erilaiset instituutiot käyttävät? |

teksteihin saisimme selville, onko tämä teksti tyypillinen lajinsa edustaja. Huomaisimme todennäköisesti, että tämän esimerkin lisäksi muutkin vastaavat tekstit on kirjoitettu asiatyylillä tapahtumia raportoiden, ilman että kirjoittaja suoraan ilmaisee mielipidettään, tuo itseään esiin tai suuremmin selittää aiheen taustoja. Nopea vertailu muutamaan muuhun kyselytuntitiedotteeseen kuitenkin osoittaa, että tämä tiedote on lyhempi kuin useat muut lajin edustajat, ja on ehkä yllättävää, että tekstissä ei mainita alkuperäisen suullisen kysymyksen esittäjän nimeä eikä myöskään jatkokysymysten esittäjiä. Tietenkin paljon laajempi vertailu olisi tarpeen, jotta tästä voitaisiin sanoa mitään varmaa.


EDUSKUNTA
RIKSDAGEN

RSS Yh

Etusivu | Kansanedustajat | Lakien säätäminen | Hallituksen valvonta | EU-asiat | Talousarvio | Kansainväli

Kaikki tiedotteet
Brysselin asiantuntijan katsaukset
Eduskuntaryhmät
Eu-asiat
Hallintovaliokunta
Hallituksen valvonta
Kansainväliset asiat
Kirjasto
Lakien säätäminen
Lakivaliokunta
Liikenne- ja viestintävaliokunta
Maa- ja metsätalousvaliokunta
Nuorten eduskunta
Perustuslakivaliokunta
Puhemiehet
Puolustusvaliokunta
Sivistysvaliokunta
Sosiaali- ja terveysvaliokunta
Suuri valiokunta
Talousarvio
Talousvaliokunta

Tiedotteet » Pienituloisten eläkeläisten ahdinko esillä kyselytunnilla

Pienituloisten eläkeläisten ahdinko esillä kyselytunnilla

Pienituloisten eläkeläisten toimeentulo oli vahvasti esillä eduskunnan kyselytunnilla torstaina 21. syyskuuta. Hallitusta kritisoitiin siitä, että se leikkaa jo valmiiksi kaikkein heikoimmassa asemassa olevilta ihmisiltä.

Kansanedustajilla oli huoli siitä, että lääkekorvausten leikkaukset ja terveydenhuollon asiakasmaksujen 30 prosentin korotus koettelevat erityisesti ikääntyviä ihmisiä. Käteen jäävä eläke ei enää riittää elämiseen, kyselytunnilla todettiin.

Sosiaali- ja terveysministeri Pirkko Mattilalta (si.) kysyttiin, onko oikein, että budjetti antaa jopa 400 euroa varakkaille, kun köyhimmiltä kansaneläkeläisiltä otetaan pois.

"Meillä on olemassa tukijärjestelmä asumistukeen ja toimeentuloon", Mattila sanoi. Hän totesi myös, että kunnat voivat itse päättää, perivätkö ne esimerkiksi terveyskeskusten asiakasmaksuja vai eivät.

Muita kyselytunnin aiheita olivat viittomakielen tulkkauspalvelut, huumeiden käyttö ja kielikokeilut.

Kyselytunnin keskustelun voi lukea kokonaisuudessaan täysistunnon päätyttyä eduskunnan pöytäkirjasta PTK 91/2017 vp tai katsoa jälkikäteen verkkolähetyksen tallenteena.

Täysistunto

Julkaistu 21.9.2017 20:05
Muokattu 21.9.2017 20:05

Esimerkki 1. Kuvakaappaus eduskunnan internetsivuilta, 21.9.2017 julkaistu tiedote.

Tekstiä voitaisiin tarkastella myös osana intertekstuaalista ketjua (Fairclough 2003), johon kuuluvat ainakin hallituksen esitys kansaneläkeindeksin jäädyttämisestä vuoden 2017 tasolle vuodeksi 2018 (Hallituksen esitys HE1232017 vp), jota käsiteltiin samana päivänä, itse kyselytunti videotallenteineen, Helsingin Sanomien artikkeli, johon kysyjä puheenvuorossaan viittasi, pöytäkirja, johon puheenvuorot on kirjattu sanasanaisesti sekä tämä tiedote. Tutkimalla tätä tekstiketjua saisimme esimerkiksi selville, mitä käsiteltyjä asioita ja kommentteja tiedotteeseen valittiin ja mitä siitä jätettiin pois. Tämä näkökulma avaa kirjoittajan prioriteetteja: kaikilla teksteillä tehdään tiettyjä asioita erityisen merkityksellisiksi ja jätetään toisia vähemmälle huomiolle tai kokonaan mainitsematta. Näitä valintoja on usein vaikea yksiselitteisesti tutkia, mutta tämän intertekstuaalisen ketjun tapauksessa voimme verrata keskustelusta tehtyä tiedotetta alkuperäiseen keskusteluun, koska molemmat ovat julkisesti saatavilla. Jo tiedotteesta näemme, että sen kirjoittaja keskittyy keskusteluun eläkeläisten toimeentulosta ja mainitsee muut kyselytunnin aiheet vain ohimennen. Pöytäkirjasta selviää lisäksi, että 21 eri henkilöä käytti eläkekeskustelussa yhteensä 38 puheenvuoroa – mutta tiedotteeseen on päätynyt vain sosiaali- ja terveysministeri Pirkko Mattilan nimi.

Tämän tyyppisestä tekstistä voisi myös olla kiinnostavaa tarkastella, miten tekstissä tuodaan esiin eri toimijoita ja kuinka aktiivisina heidät esitetään. Tämä tiedote esimerkiksi asettaa vastakkain ministeri Mattilan ja nimeämättömän joukon kansanedustajia. Epätarkkuutta toimijoiden suhteen lisää passiivin (*kritisoitiin, todettiin, kysyttiin*) ja muiden persoonaviittauksen osalta epäselvien ilmausten käyttö (*toimeentulo oli vahvasti esillä, Kansanedustajilla oli huoli*). Usein analyysiprosessissa voi olla hyödyllistä miettiä, miten muuten saman asian olisi voinut ilmaista (ks. Kielitaju kielentutkijan työkaluna tk.): esimerkiksi tässä olisi voinut myös mainita, ketkä tai minkä puolueen edustajat hallitusta kritisoivat. Tällaisen tiedon puuttuminen voi luoda mielikuvaa, että oppositio olisi ollut yksimielinen asian suhteen tai että Mattila olisi vastannut kritiikkiin tyhjentävästi.

Lisäksi kiinnostavaa voisi olla analysoida tekstin sisäistä koherenssia ja sitä, miten asioita siinä yhdistetään ja rinnastetaan. Esimerkiksi

tiedotteessa mainittu kysymys ja vastaus eivät tunnu kuuluvan yhteen (eivätkä pöytäkirjan mukaan kuulukaan). Itse asiassa ministeri Mattila sanoi, useita puheenvuoroja myöhemmin: "Meillä on edelleen olemassa tukijärjestelmä asumistukeen, toimeentuloon, myöskin lääkkeisiin, ja niin kuin tässä kuultiin, tämä asiakasmaksulaki myöskin turvaa sen, että se voidaan jättää perimättä vähävaraisuuden vuoksi".

Erityisesti kriittisesti suuntautuneessa diskurssianalyysissa ollaan usein kiinnostuneita tekstin sisältämistä taustaoletuksista; esimerkiksi siitä, mitkä asiat nähdään normaaleina, hyvinä ja arvokkaina. Tässä tiedotteessa verbi *kritisoitiin* viestii siitä, että kritiikin esittäjät pitävät heikoimmassa asemassa olevilta leikkaamista huonona tai moraalittomana ratkaisuna. Kansanedustajien *huoli* viittaa inhimillisiin tunteisiin pienituloisia eläkeläisiä kohtaan, joita *koetellaan*. Toisaalta myös tiedotteen otsikossa on tunnepitoinen sana *ahdinko*, mikä vihjaa, että kirjoittaja ehkä hyväksyy saman taustaoletuksen. Jälleen voidaan miettiä, millä muilla tavoin asian olisi voinut halutessaan muotoilla. Sanat *huoli* ja *ahdinko* avaavat merkityksiensä kautta kiinnostavasti näkyville myös emootioiden roolin poliittisessa keskustelussa. Kysymys, onko *oikein*, että varakkaille annetaan ja köyhimmiltä otetaan pois, puolestaan tuo moraalisen pohdinnan etualalle. Kysymys kuulostaa ehkä niin sanotulta retoriselta kysymykseltä, mutta kyselytunnin yhteydessä kysymys oli sinänsä aito, että ministerin odotettiin vastaavan siihen ja hän vastasikin.

## 3.4. Esimerkki suulliseen vuorovaikutusaineistoon perustuvasta diskurssianalyyttisesta tulkinnasta

Toinen esimerkkimme on keskustelukatkelma videoidusta keskustelutilanteesta keskisuuressa, noin 500 henkilöä työllistävässä rahoitus- ja myyntiorganisaatiossa (ks. esimerkki 2). Katkelmassa keskustelevat yrityksen toimitusjohtaja (Esa) ja markkinointijohtaja (Jan). Meneillään olevaa keskustelua nimitettiin kyseisessä yrityksessä johtamiskeskusteluksi, mikä tarkoitti sitä, että toimitusjohtaja kävi jokaisen ylempään johtoon kuuluvan kanssa läpi keskeisiä asioita yrityksen uudesta strategiasta. Ennen keskustelua johtajat olivat täyttäneet kyselylomakkeen,

jossa he olivat numeerisesti ja kirjallisesti arvioineet sekä yrityksen että myös omien henkilökohtaisten tavoitteidensa saavuttamista ja tavoitteisiin liittyviä haasteita.

Alkuperäinen syy tämän aineiston keräämiseen oli tutkia, miten johtoasemissa olevat ihmiset puhuvat strategiasta. Mutta toisaalta aitoja kielenkäyttötilanteita sisältäville aineistoille on tyypillistä, että ne taipuvat moniin tarkoituksiin. Kyseisiä keskusteluja voidaan siis tulkita strategiakontekstia vasten ja asettaa aineistolle kysymyksiä, kuten miten strategiasta puhutaan. Kysymykset voisivat kielenkäytön analyysin tasolla tarkentua esimerkiksi siihen, millaisia metaforia strategiasta käytetään tai millaisissa toimija- tai subjektiasemissa se esiintyy (esimerkki tästä ks. Pälli 2018). Samaa keskustelua voisi toki katsoa myös hieman erilaisesta näkökulmasta: sen voisi nähdä ilmentävän esimerkiksi johtamista tilanteisena toimintana eli sitä, miten keskustelussa ”johdetaan”. Voisi myös kysyä, mitä johtaminen on, millaisia merkityksiä puhujat siihen liittävät ja millaisia kielen keinoja he näiden merkitysten tuottamiseen käyttävät. Lisäksi tutkijaa voisi kiinnostaa esimerkiksi työpaikkakulttuuri, joka voitaisiin edelleen yhdistää diskurssianalyyttiseen keskusteluun vaikkapa naisten ja miesten diskursiivisista johtamistavoista (Holmes, Schnurr & Marra 2007). Konteksteja voisi olla monia muitakin diskurssianalyyttisen mielikuvituksen – kyvyn yhdistää kielenkäyttö sosiaaliseen toimintaan tai johonkin sosiaalisen toiminnan kontekstiin – tarjoamissa rajoissa.

Myös suullisten vuorovaikutusaineistojen diskurssianalyysi voi perustua enemmän kielelliseen havaintoon ja jättää vähemmälle edellä mainitun kaltaiset sosiaaliset kontekstit. Diskurssianalyytikko voi huomata tällaisesta aineistosta jonkin kiinnostavan kielenkäyttöön liittyvän ilmiön, jonka kiinnostavuus ei välttämättä ole juuri ollenkaan tekemisissä esimerkiksi keskustelun puheenaiheen kanssa. Yhtenä mahdollisena tulkintasuuntana voisi esittää vaikkapa seuraavan havainnon, joka voisi olla mielenkiintoinen myös muissa konteksteissa: Esimerkissä rivillä 34 Jan sanoo *täs on halunnu reagoida- tai olen reagoinu* tuoden näin esiin oman tietoisuutensa vastuullisesta tekijyydestä. Kolmannen persoonan muotoilu *on halunnu reagoida* sisältää implisiittisesti puhujan mutta samalla yleistää hänet osaksi laajempaa joukkoa. Tekijyyden ilmaisuun

liittyy kuitenkin itsekorjaus *olen reagoinu,* jonka merkkinä voidaan nähdä kesken jäävä *reagoida*-sana ja sen jälkeinen *tai*. Nimenomaan itsekorjauksen avulla Jan säilyttää ennen korjausta sanotun relevanssin (vrt. Sorjonen & Laakso 2005) ja toisaalta itse alleviivaa tekijyyden, omakohtaisuuden, olevan merkityksellinen kielellinen valinta tässä tilanteessa. Riveillä 36–38 Jan jatkaa kommentilla *toi meni itse asiassa tunnepuolelle*. Tässäkin voidaan nähdä tekijyyden tietynlainen etäännyttäminen, sillä valittu kielellinen kuvaustapa korostaa puheena olleiden asioiden ”menneen” jättäen samalla tekijän vastuun taka-alalle. Tulkintaa tukee Janin kehollinen toiminta. Hän suuntaa puhuessaan katseensa paperiin ja ohi keskustelukumppaninsa ja ristii kätensä niin kutsuttuun torjunta-asentoon rinnalleen (rivi 37).

Diskurssianalyysia edustaisi myös enemmän vuorovaikutustoimintaan perustuva tulkinta siitä, miten organisatorista asemaa rakennetaan ja pönkitetään keskustelussa kielen ja vuorovaikutuksen keinoin. Analyysissa voisi katsoa systemaattisesti sitä, kuka avaa ja päättää puheenaiheita eli topiikkeja ja millä keinoin. Tällaista analyysia tehtäessä ollaan samoilla vesillä keskusteluntutkimuksellisen tulkinnan kanssa. Eroksi voidaan nostaa kuitenkin diskurssianalyysille tyypillinen vaatimus myös muun kuin keskustelun kontekstin olennaisesta merkityksestä: analyysin tekijä haluaisi ehkä tutkimuksellaan kertoa nimenomaan siitä, miten topiikkien avaaminen, käsittely ja lopettaminen liittyy esimerkiksi johtamistyyliin (Svennevig 2011). Diskurssianalyysille on siis leimallista pyrkiä sosiaalisen makrotason tulkintoihin, kun keskustelunanalyysi kohdentuu tiukemmin juuri keskustelun itsensä kontekstiin.

Esimerkki 2. Kahden välistä johtamiskeskustelua organisaatiossa X. Toimitusjohtaja (Esa) ja markkinointijohtaja (Jan).

Litterointimerkinnät: - kesken jäävä ilmaus; [ ] päällekkäispuhunta; *no* alleviivaus merkitsee painotusta; (.) lyhyt tauko; (1.0) tauko sekunneissa; = kaksi puhunnosta liittyy toisiinsa tauotta; (( )) litteroijan kommentit.

01 Esa: Onko tota (.) koulutus- (.) ni onks sitä ollu riittävän paljo
02 tässä (.) nyt lähivuosina vai-
03 Jan: Mm- mehän käytiin (.) siis mä kävin nää peruskurssit (.) sit on
04 muutama preppaus ollu jostaki tuotteesta. Ja sitte oli tää (.) Mörkö
05 tää- mä en muista enää nimeä ne sano rumasti Möröks sitä yks
06 konsultti (.) jonka Miettinen (.) niinku laitto meidät laittaan
07 myyntikoulutusta=se oli hyvin erikoinen rupeama se ajo jotain (.)
08 käärmettä pulloon yritti ajaa meille. Se oli hyvin erikoinen. Sit oli
09 tää- (.) sit oli myyntikoulutusta oli tää jalkapallo Anssi T. Salo
10 vai mikä se on. Et ei sitä kauheesti oo ollu kyllä. Ei oo ollu et
11 kyllähän meidän täytyy nyt jatkos ni koulutuskalenteriin määritellä
12 että siel on ((sanoja poistettu)) mis meil on oikeasti ongelmia (.)
13 mut ei sitä hurjasti oo ollu kyllä. Aina vaan vähentyny tässä viime
14 vuosina. Et siinä on ollu joku ketunhäntä kainalos ku sitä on tehty.
15 Esa: Niin voi olla joo. Joo. (4.0) ((molemmat merkitsevät jotain
16 papereihin)) Sitte kehittämis- (1.0) työ.
17 Jan: No tää nyt kyllä on vähä ehkä liiankin raaka tää ykkönen mut tässä
18 mä olin niin tuohduksissa sillon kun (.) mä oikeesti- (.) no tähän

kytkeytyy nyt se huoli tästä talouden tilanteesta ihan oikeesti mä
mietin et mikä on se ((yrityksen nimi)) olemassa olon oikeutus
((sanoja poistettu). Ja sitte ku nää puuttuvat työkalut jotka ihan
oikeesti niin nehän kaikilla kilpailijoilla on jo. Ni se täs
aiheuttaa sitä minkä takia mä nyt oon sen [tänne alle-
Esa: [Mikä se työkalu on?
Jan: Työkalut on (.) se liittyy tähän tekniikkaan=ne on nää e-laskut
sähköset palvelut, lakisääteiset palkkailmoitukset,
((sana poistettu)) tuote. ((Listaa puhuessaan sormilla, katse ohi))
Esa: Niin nämä. Joo.
Jan: Niin nehän on niinku ihan, ne on semmoset mä nään niinku (.) tosi
pahana juttuna et niit ei oo. Et jos mul ois rahaa niin mä ostaisin
ne äkkiä tälle talolle. Ne on niinku se- sen suuruusluokan asioita
mun mielestä.
Esa: Joo. (.) Joo. (1.0)
Jan: Sen takia täs on halunnu reagoida- tai olen reagoinu noin vahvasti.
Esa: Joo mutta sinänsä tää on must (.) oikeast[aan-
Jan: [Toi meni itseasias
tunnepuolelle jo ((nauraa))
Esa: Mm toi- menettelyhomma niin no mä oon kertonu tästä että (.) nää
on todellisuudessa (.) näiden tarkotus on toimii Normandian
maihinnousun savuverhona.
Jan: Niin.

42 Esa: Sit sä laitat jonkun numeron vaan sinne niin se oleellinen asia
43 on siellä alla oleva ((näyttää elein tekstilaatikon ja hymyilee))
44 Jan: Niin teksti, kyllä.
45 Esa: Et se on niinku tää että mä oon nähny kuinka, esimiehelle josta
46 kaikki sanoo että sille ei kukaan anna uskalla ykköstä antaa
47 vaikka se on ykkösen arvonen=mä olen nähny ku sille annetaan se
48 ykkönen. Koska siellä alla lukee että mitä pitäisi tapahtua
49 että esimiestyöni olisi paremmin. Siin lukee alla sitte että (.)
50 minusta meidän pitäisi tavata vähintään kerran kuukaudessa
51 keskustelemassa minun työtilanteesta ja niin edelleen. Mä en oo
52 ainakaan nähny ketään joka ois ottanu herneen nenäänsä sitte siitä.
53 Jan: Niin (.) se asia on siinä.
54 Esa: Siis se ykkönen häviää sinne ja se on- miettii vaan että (.) ei ne
55 nyt ihan tolkuttomat noi vaateet oo. Ihan (.) täysjärkisii (.)
56 no mutta mä en sinänsä kiinnitä (.) mä ite oon tämmönen mä annan
57 ykköstä ja kakkosta.
58 Jan: Joo no tässä nyt on halunnu pysähtyä tohon oikeesti.

Analyysissa voisi tarkastella myös sitä, miten sosiaalista ongelmallisuutta käsitellään. Esimerkissä sosiaalisesti ongelmallisena voitaisiin nähdä Janin antama heikko numeerinen arviointi, joka koskee Esan rivillä 16 avaamaa topiikkia kehittämistyö. Ongelmallisuudesta kertoo esimerkiksi psykologinen selonteko (engl. *account*) *olin niin tuohduksissa* (rivi 18) ja sitä kehystävä arviointi (*vähä ehkä liiankin raaka tää ykkönen*). Sosiaalisesta ongelmallisuudesta ja tietynlaisen sosiaalisen tunnustuksen

tehneen Janin ”kasvojen” varjelemisesta kertoo puolestaan Esan (r. 54–57) todistelu siitä, että arvioinnin numerot eivät ole niin tärkeitä kuin sanallinen arviointi. Tässäkin yhteydessä keskustelussa kiinnostavasti psykologisoidaan eli selitetään arvioinnin tapa henkilökohtaisena ominaisuutena (r. 56–57: *mä ite oon tämmönen mä annan ykköstä ja kakkosta*).

Esimerkkiä voitaisiin analysoida myös multimodaalisesti eli käyttämällä videokuvatun aineiston suomia mahdollisuuksia yhdistää keskustelun verbaalin kielen analyysi keskustelijoiden keholliseen vuorovaikutukseen (esim. Knoblauch, Schnettler & Raab 2006). Tässä artikkelissa emme anonymiteettilupauksen vuoksi voi näyttää videotallennetta tai kuvakaappauksia tallenteesta, mutta teemme pari esimerkinomaista huomiota. Diskurssianalyyttisen tulkinnan kannalta kiintoisaa voisi olla esimerkiksi, kuinka papereita käytetään apuna ja välineinä vuorovaikutuksessa (läpi koko esimerkin puhujat sormeilevat ja vilkuilevat heidän välissään pöydällä olevia papereita), kuinka rytmiset eleet esiintyvät painottamisen ja listan tuottamisen välineinä (r. 27) tai kuinka ikoniset eleet voivat ohjata ja täsmentää merkityksiä (r. 43).

Vaikka useimmissa tapauksissa tutkijalla on ainakin alustavasti tiedossaan joko omaan havaintoon tai johonkin muuhun lähteeseen perustuen aihepiiri, jonka kielellistä ja vuorovaikutuksellista rakentumista hän aineistosta katsoo, ilmiön tarkempi määrittely syntyy usein analyysin kautta. Tarkempi pureutuminen kontekstuaaliseen kielenkäyttöön täsmentää analyysin kysymyksenasetteluja. Jos esimerkkinämme toimivaa keskustelukatkelmaa luettaisiin aineistona esimiestyöskentelystä ja vallan rakentumisesta keskustelussa, voisi analyysin tekijä paikantaa lukuisia eritasoisia kielellisiä (tai multimodaalisia) rakenteita ja niiden funktioita, joihin analyysi voisi perustua. Rakenteisiin ja niiden kautta funktioihin liittyisi jo aiemmin mainitsemamme tekijyyden esittäminen ja persoonamuotoisuus. Esimerkin yhteydessä voitaisiin puhua johtamis-, tavoite- tai kehityskeskustelutekstilajista (tai sosiaalisesta toimintatyypistä), johon liittyisi esimerkiksi keskustelun rakenteellinen kysymys–vastaus-muotoisuus ja puhujien esiin tuoma ymmärrys rooleistaan keskustelussa. Myös sekä verbaaliset että keholliset metaforat ja analogiat ja nimeämisen tavat ovat nekin esimerkissä hyvin kiintoisia.

# 4. Yhteenveto

Diskurssianalyysin dialoginen ja sosiaalinen, vuorovaikutusta korostava kielikäsitys ja konstruktionistinen perusta katsoa kieltä todellisuuden sosiaalisen rakentumisen välineenä ovat diskurssianalyyttisen metodin puitteet. Kielenkäytön mikrotason suhteen diskurssianalyysin metodiset ohjenuorat ovat kuitenkin väljät. Ei ole olemassa yhtä tiettyä analyysivälinettä tai esimerkiksi kielenilmiöiden listaa, jonka sisältämät kohdat aineistosta analysoimalla saisi tehtyä oikeanlaisen diskurssianalyysin.

Kielenkäyttöön ja rakenteeseen liittyviä analyysivälineitä kuitenkin riittää, ja tässä mielessä diskurssianalyysin metodinen maisema on sangen rikas. Tässä artikkelissa olemme kuvanneet välineistä muutamia ja avanneet lyhyesti kahden hyvin erityyppisen esimerkin kautta sitä, millä tavoin kielenkäytöstä havainnoitujen kielellisten tai vuorovaikutuksellisten ilmiöiden voi nähdä antavan perusteita sosiaaliselle tulkinnalle. Avainkysymys diskurssianalyysissa onkin juuri se, miten kielenkäytöstä tehdään sosiaalista tulkintaa.

Sosiaalinen tulkinta vaatii lingvististä ja nimenomaan diskurssianalyyttista mielikuvitusta – jolle läheinen käsite voisi olla myös *kielitaju* (ks. Kielitaju kielentutkijan työkaluna tk.). Joka tapauksessa meillä kaikilla on valtava kokemuksellinen varanto erilaisia huomioita kielenkäytön ja kielen rakenteiden toistuvuuksista sekä tulkinnoista, joita olemme tehneet ja joista olemme kuulleet liittyen kielenkäytön tapoihin erilaisissa tilanteissa. Kun puhumme diskurssianalyyttisesta mielikuvituksesta, lainaamme C. Wright Millsin (2015 [1959]) ajatusta sosiologisesta mielikuvituksesta, jolla Mills viittasi tietoisuuteen henkilökohtaisen kokemuksen ja sosiaalisten rakenteiden suhteesta. Vastaavanlainen käsite on *sosiologinen ambivalenssi* (Merton 1976), jolla tarkoitetaan sitä, että sosiologi on aina osa sitä, mitä tutkii. Muutettavat muuttaen: diskurssianalyytikko tutkii asioita, jotka ovat hänen omassa kokemuspiirissään, ja diskurssianalyytikolla on oltava mielikuvitusta nähdä yhteyksiä kielenkäytön ja sosiaalisen toiminnan välillä. Kenttä on laaja, ja monet asiat ja ilmiöt voivat olla tutkimuksen kohteena. Itse asiassa diskurssianalyyttinen mielikuvitus yhdistää nimenomaan sosiologisen mielikuvituksen

ja lingvistisen mielikuvituksen; diskurssianalyytikko oivaltaa yhteyksiä sosiaalisten rakenteiden ja kielenkäytön rakenteiden tai niiden käytön välillä.

**Aiheesta lisää:**


Dijk, Teun van: "Discourse and knowledge". [verkkoaineisto]. [viitattu 18.10.2020]. Saatavissa: https://youtu.be/sxfc-WJRKEM.

Johnstone, Barbara. 2008. *Discourse Analysis.* (2. p.). Malden (MA): Blackwell.

Jokinen, Arja, Juhila, Kirsi & Suoninen, Eero. 2016. *Diskurssianalyysi*. Tampere: Vastapaino.

Jones, Rodney. 2012. *Discourse Analysis: A Resource Book for Students*. Abingdon: Routledge.

Pietikäinen, Sari & Mäntynen, Anne. 2009. *Kurssi kohti diskurssia*. Tampere: Vastapaino.

Potter, Jonathan: "Discourse & cognition: Issues of interaction, emotion and mental ontology". [verkkoaineisto]. [viitattu 18.10.2020]. Saatavissa: https://youtu.be/jChg4PQvfjA.


## VIITTEET

1 Esimerkki tällaisesta on organisaatiotutkimuksessa virinnyt kiinnostus strategiatyön käytäntöihin. Menetelminä on hyödynnetty esimerkiksi kriittistä diskurssianalyysia, keskusteluntutkimusta, etnografiaa ja toimintatutkimusta. Yhteinen ilmiö, jota tarkastellaan, on kuitenkin nimenomaan strategiatyö. (Esim. Sorsa, Pälli, Vaara & Peltola 2010.)

## KIRJALLISUUS


Angermuller, Johannes, Mainguenau, Dominique & Wodak, Ruth. 2014. The discourse studies reader: An introduction. Julkaisussa: Angermuller, Johannes, Maingueneau, Dominique & Wodak, Ruth (toim.) *The Discourse Studies Reader: Main Currents in Theory and Analysis*. Amsterdam: Benjamins, 1–14.

Austin, John L. 1962. *How to Do Things with Words*. Oxford: Oxford University Press.

Bahtin [Bakhtin], Mihail. 1981. *The Dialogic Imagination: Four Essays*. Toim. Michael Holquist, englanniksi kääntäneet Caryl Emerson & Michael Holquist. Austin: University of Texas Press.

Bahtin [Bakhtin], Mihail. 1986 [1953]. The problem of speech genres. Englanniksi kääntänyt Vern McGee. Julkaisussa: Emerson, Caryl & Holquist, Michael (toim.) *M. M. Bakhtin, Speech Genres and Other Late Essays*. Austin: University of Texas Press, 60–102.

Baker, Paul, Gabrielatos, Costas, KhosraviNik, Majid, Krzyzanowski, Michał, McEnery, Tony & Wodak, Ruth. 2008. A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society* 19, 273–306.

Barthes, Roland. 1993. *Tekijän kuolema, tekstin syntymä*. Suom. Lea Rojola ja Pirjo Thorel. Tampere: Vastapaino.

Bazerman, Charles. 1997. Discursively structured activities. *Mind, Culture, and Activity* 4, 296–308.

Benwell, Bethan & Stokoe, Elizabeth. 2006. *Discourse and Identity*. Edinburgh: Edinburgh University Press.

Berger, Peter & Luckmann, Thomas. 1966. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. Garden City (NY): Anchor Books.

Blommaert, Jan. 2009. On Hymes. Introduction. *Text & Talk* 29, 241–243.

Brown, Gillian & Yule, George. 1983. *Discourse Analysis*. Cambridge: Cambridge University Press.

Brown, Penelope & Levinson, Stephen. 1978. Universals in language usage: Politeness phenomena. Julkaisussa: Goody, Esther N. (toim.) *Questions and Politeness: Strategies in Social Interaction*. Cambridge: Cambridge University Press, 56–310.

Burr, Vivien. 2015 [1995]. *Social Constructionism*. (3. p.). London: Routledge.

Daneš, František. 1974. Functional sentence perspective and the organisation of the text. Julkaisussa: Daneš, František (toim.) *Papers on Functional Sentence Perspective*. Prague: Academia, 106–128.

Davies, Bronwyn & Harré, Rom. 1990. Positioning: the discursive production of selves. *Journal for the Theory of Social Behavior* 20, 43–63.

De Fina, Anna & Georgakapoulou, Alexandra (toim.). 2015. *The Handbook of Narrative Analysis*. Chichester: John Wiley.

Devitt, Amy. 1991. Intertextuality in tax accounting: generic, referential, and functional. Julkaisussa: Bazerman, Charles & Paradis, James (toim.) *Textual Dynamics of the Professions: Historical and Contemporary Studies of Writing in Professional Communities*. Madison: The University of Wisconsin Press, 336–357.

Dijk, Teun van. 1972. *Some Aspects of Text Grammars: A Study in Theoretical Linguistics and Poetics*. Mouton: The Hague & Paris.

Dijk, Teun van. 1977. Semantic macro-structures and knowledge frames in discourse comprehension. Julkaisussa: Just, Marcel A. & Carpenter, Patricia A. (toim.) *Cognitive Processes in Comprehension*. Hillsdale (NJ): Lawrence Erlbaum, 3–32.

Dijk, Teun van. 1997. Introduction: The study of discourse. Julkaisussa: Dijk, Teun van (toim.) *Discourse Studies: A Multidisciplinary Introduction*. London: SAGE, 1–7.

Edwards, Derek & Potter, Jonathan. 1992. *Discursive Psychology*. London: SAGE.

Fairclough, Norman. 1992. *Discourse and Social Change*. Cambridge: Polity Press.

Fairclough, Norman. 2001 [1989]. *Language and Power.* (2. p.). London: Longman.

Fairclough, Norman. 2003. *Analysing Discourse: Textual Analysis for Social Research*. London: Routledge.

Fillmore, Charles. 1968. The case for case. Julkaisussa: Bach, Emmon & Harms, Robert (toim.) *Universals in Linguistic Theory*. New York (NY): Holt, Rinehart and Winston, 1–88.

Foucault, Michel. 1970. *The Order of Things: An Archeology of the Human Sciences*. London: Tavistock.

Fowler, Roger, Hodge, Bob, Kress, Günther & Trew, Tony. 1979. *Language and Control*. London: Routledge.

Frawley, William. 1992. *Linguistic Semantics*. Hillsdale (NJ): Lawrence Erlbaum.

Genette, Gerard. 1972. *Figures III*. Paris: Éditions du Seuil.

Grice, H. Paul. 1975. Logic and conversation. Julkaisussa: Cole, Peter & Morgan, Jerry (toim.) *Syntax and Semantics 3*: *Speech Acts*. New York (NY): Academic press, 41–58.

Gumperz, John. 1982. *Discourse Strategies. Studies in Interactional Sociolinguistics* 1. Cambridge: Cambridge University Press.

Halliday, Michael A. K. 1994 [1985]. *An Introduction to Functional Grammar.* (2. p.). London: Edward Arnold.

Halliday, Michael A. K. & Hasan, Ruqaiya. 1976. *Cohesion in English*. London: Longman.

Harris, Zellig. 1952. Discourse analysis. *Language* 28, 1–30.

Heikkinen, Vesa. 1999. *Ideologinen merkitys kriittisen tekstintutkimuksen teoriassa ja käytännössä*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä, Ulla & Lounela, Mikko (toim.). 2012. *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus.

Holmes, Janet, Schnurr, Stephanie & Marra, Meredith. 2007. Leadership and communication: discursive evidence of a workplace culture. *Discourse & Communication* 1, 433–451.

Hymes, Dell H. 1974. *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadephia: University of Pennsylvania Press.

Jokinen, Arja, Juhila, Kirsi & Suoninen, Eero. 1993. *Diskurssianalyysin aakkoset*. Tampere: Vastapaino.

Jokinen, Arja, Juhila, Kirsi & Suoninen, Eero. 1999. *Diskurssianalyysi liikkeessä*. Tampere: Vastapaino.

Jokinen, Arja, Juhila, Kirsi & Suoninen, Eero. 2016. *Diskurssianalyysi*. Tampere: Vastapaino.

Karvonen, Pirjo. 1995. *Oppikirjateksti toimintana*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Knoblauch, Hubert, Schnettler, Bernt & Raab, Jürgen. 2006. Video-analysis: Methodological aspects of interpretive audiovisual analysis in social research. Julkaisussa: Knoblauch, Hubert, Schnettler, Bernt, Raab, Jürgen & Soeffner, Hans-Georg (toim.) *Video-Analysis: Methodology and Methods. Qualitative Audiovisual Data Analysis in Sociology*. Berlin: Peter Lang, 9–28.

Kristeva, Julia. 1986 [1966]. Word, dialogue and novel. Julkaisussa: Moi, Toril (toim.) *The Kristeva Reader*. New York (NY): Columbia University Press, 34–62.

Leech, Geoffrey. 1974. *Semantics*. Harmondsworth: Penguin Books.

Leeuwen, Theo van. 2007. Legitimation in discourse and communication. *Discourse & Communication* 1, 91–112.

Levinson, Stephen C. 2004. Deixis. Julkaisussa: Horn, Laurence (toim.) *The Handbook of Pragmatics*. Oxford: Blackwell, 97–121.

Lillqvist, Ella. 2019. Korpusavusteinen diskurssianalyysi kuluttajatutkimuksen menetelmänä: Pikavippikeskustelun synty, nousu ja arkipäiväistyminen Suomi24-keskustelufoorumilla. *Kulutustutkimus.Nyt* 13, 5–30.

Maschler, Yuel & Schiffrin, Deborah. 2015. Discourse markers: language, meaning, and context. Julkaisussa: Tannen, Deborah, Hamilton, Heidi E. & Schiffrin, Deborah (toim.) *The Handbook of Discourse Analysis*. (2. p.). Chichester: John Wiley & Sons, Ltd, 189–221.

Mautner, Gerlinde. 2016. Checks and balances: How corpus linguistics can contribute to CDA. Julkaisussa: Wodak, Ruth & Meyer, Michael (toim.) *Methods of Critical Discourse Studies*. London: SAGE, 122–143.

Merton, Robert C. 1976. *Sociological Ambivalence and Other Essays*. New York (NY): The Free Press.

Mills, C. Wright. 2015 [1959]. *Sosiologinen mielikuvitus*. Suom. Antti Karisto, Esa Konttinen, Pentti Takala & Hannu Uusitalo. Helsinki: Gaudeamus.

Mäntynen, Anne, Shore, Susanna & Solin, Anna. 2006. *Genre – tekstilaji*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Nuyts, Jan. 2005. The expression of modality. Julkaisussa: William Frawley (toim.) *The Expression of Cognitive Categories 1*. Berlin: Mouton de Gruyter, 1–26.

Paltridge, Brian. 2012. *Discourse Analysis. An Introduction*. (2. p.). London: Continuum.

Pietikäinen, Sari & Mäntynen, Anne. 2009. *Kurssi kohti diskurssia*. Tampere: Vastapaino.

Potter, Jonathan. 1996. *Representing Reality. Discourse, Rhetoric and Social Construction*. London: SAGE.

Potter, Jonathan & Wetherell, Margaret. 1987. *Discourse and Social Psychology: Beyond Attitudes and Behaviour*. London: SAGE.

Pälli, Pekka. 2003. *Ihmisryhmä diskurssissa ja diskurssina*. Acta Universitatis Tamperensis 910. Tampere: Tampere University Press.

Pälli, Pekka. 2018. Ascribing materiality and agency to strategy in interaction: A language-based approach to the material agency of strategy. *Long Range Planning* 51, 436–450.

Schiffrin, Deborah. 1994. *Approaches to Discourse*. London: Blackwell.

Searle, John R. 1975. Indirect speech acts. Julkaisussa: Cole, Peter & Morgan, Jerry (toim.) *Syntax and Semantics, 3: Speech Acts*. New York (NY): Academic Press, 59–82.

Sinclair, John McHardy & Coulthard, Malcolm. 1975. *Towards an Analysis of Discourse: the English Used by Teachers and Pupils*. London: Oxford University Press.

Sorjonen, Marja-Leena & Laakso, Johanna. 2005. Katko vai eiku? Itsekorjauksen aloitustavat ja vuorovaikutustehtävät. *Virittäjä* 109, 244–271.
Sorsa, Virpi, Pälli, Pekka, Vaara, Eero & Peltola, Katja. 2010. *Strategia mahdollisuutena ja rajoitteena kuntaorganisaatiossa. Kielestä, kommunikaatiosta ja vallasta*. Research Reports 70. Helsinki: Hanken School of Economics.
Svennevig, Jan. 2011. Leadership style in managers' feedback in meetings. Julkaisussa: Angouri, Jo & Marra, Meredith (toim.) *Constructing Identities at Work*. Basingstoke: Palgrave Macmillan, 17–39.
Tieteen termipankki. 2019. [verkkoaineisto]. [viitattu 12.2.2020]. Saatavissa: https://tieteentermipankki.fi.
Wiggins, Sally & Potter, Jonathan. 2008. Discursive psychology. Julkaisussa: Willig, Carla & Rogers, Wendy (toim.) *The SAGE Handbook of Qualitative Research in Psychology*. London: SAGE, 73–90.
Wittgenstein, Ludwig. 2001 [1953]. *Filosofisia tutkimuksia*. Suom. Heikki Nyman. Helsinki: WSOY.
Wodak, Ruth & Meyer, Michael. 2016 [2001]. Critical discourse analysis: History, agenda, theory and methodology. Julkaisussa: Wodak, Ruth & Meyer, Michael (toim.) *Methods of Critical Discourse Studies*. (3. p.). London: SAGE, 1–22.

# Luku 7 Laadullinen aineistopohjainen kielentutkimus

*Milla Luodonpää-Manni*
https://orcid.org/0000-0002-1646-4462

*Krista Ojutkangas*
https://orcid.org/0000-0002-4479-4637

**Mikä?**

Laadullinen aineistopohjainen kielentutkimus on aitojen kielenkäyttöaineistojen syvällistä tarkastelua. Vaikka kielellisten aineistojen tallennuksessa, käsittelyssä ja analyysissa hyödynnetään usein digitaalisia apuvälineitä, tutkijan tulkinnalla ja käsityönä tekemällä analyysilla on menetelmässä keskeinen asema, sillä laadullisella aineistopohjaisella kielentutkimuksella tavoitellaan tutkimuskohteen kokonaisvaltaista ymmärtämistä.

**Katso myös:**

Johdannon luku 1. Tutkimuksen käytäntö

Luku 9 Korpusaineistot

Luku 10 Määrällinen korpuslingvistiikka

Luku 11 Historiallinen korpuslingvistiikka

# 1. Johdanto: mitä on laadullinen aineistopohjainen kielentutkimus?

Suomalaisessa kielentutkimuksessa on hyvin tavallista, että kielellisistä ilmiöistä hankitaan tietoa tarkastelemalla tutkimuskohdetta laadullisesti aidoissa kielenkäyttöaineistoissa. Laadullista aineistopohjaista kielentutkimusta tekevää opiskelijaa tai nuorta tutkijaa saattaa kuitenkin hämmentää kysymys tutkimuksen menetelmästä. Jos käytössä ei ole jotakin vakiintunutta, erikseen nimettyä menetelmää, kuten keskustelunanalyysia, tarkoittaako tämä sitä, että tutkimuksessa ei ole mitään menetelmää?

Ajatus siitä, että tieteellisessä tutkimuksessa ei noudatettaisi mitään systemaattista menettelytapaa, jolla aineistosta hankitaan tietoa, on loogisesti ja eettisesti mahdoton. Väistämättä on niin, että aineistoa tarkastellessaan tutkija jäsentää ja järjestää tutkimuskohdetta erilaisiin luokkiin havaintojensa ja ennakkotietojensa perusteella. Tästä menetelmästä käytetään kielentutkimuksessa toisinaan nimitystä *perinteinen lingvistinen analyysi*. Käytännössä järjestämisen apuna voidaan käyttää lähes mitä tahansa systemaattista menettelytapaa, esimerkiksi aineistoesimerkkejä sisältäviä pahvilappuja, jotka järjestetään kenkälaatikkoon tietyn periaatteen mukaisesti, tai värikynämenetelmää, jossa samaan luokkaan kuuluvat tapaukset merkitään yhtenäisellä värikoodilla. Nykyisin aineiston järjestämisessä hyödynnetään yhä useammin digitaalisia apuvälineitä. Riippumatta luokittelun käytännön toteutuksesta aineiston järjestämisen taustalla olevat periaatteet on tärkeä tunnistaa ja kuvata tutkimuksen menetelmäosiossa.

Keskitymme tässä artikkelissa kuvaamaan sellaista laadullista aineistopohjaista kielentutkimusta, joka kohdistuu digitaalisista aineistoista poimimalla koostettuun tutkimusaineistoon ja joka toteutetaan tietokantaohjelman avulla.[1] Tutkimusasetelmasta johtuen käytämme tässä artikkelissa selkeyden vuoksi sanaa *tutkimusaineisto* tarkoittamaan tutkijan itselleen koostamaa esiintymien joukkoa, jonka perusteella tutkimus tehdään. Tässä artikkelissa esiteltyjä periaatteita voidaan kuitenkin hyvin soveltaa kaikenlaiseen aineistopohjaiseen laadulliseen analyysiin.

Laadullisella aineistopohjaisella kielentutkimuksella ymmärrämme tässä aitojen kielenkäyttöaineistojen syvällistä tarkastelua, jonka tavoitteena on tutkimuskohteen kokonaisvaltainen ymmärtäminen (ks. myös Hasko 2012). Digitaalisessa muodossa olevia aineistoja on mahdollista tallentaa, käsitellä ja analysoida tietokoneavusteisesti. Kielentutkimuksen materiaalina ei kuitenkaan tarvitse käyttää pelkästään valmiiksi annotoituja massa-aineistoja, joita on saatavilla eri kielille (ks. Korpusaineistot tässä kirjassa), vaan tutkijalla on vapaus koota oma tutkimusaineistonsa ja käyttää myös ei-digitaalisessa muodossa olevaa materiaalia sen koostamiseen. Monista hyvistä puolistaan huolimatta valmiit massa-aineistot eivät pysty vastaamaan kaikkiin kielentutkimuksen tarpeisiin. Laadulliselle aineistopohjaiselle kielentutkimukselle onkin ominaista määrällistä korpustutkimusta pienempien tutkimusaineistojen käyttö (ks. Määrällinen korpuslingvistiikka tk.). Pienemmän tutkimusaineiston käyttö mahdollistaa myös tutkimuskohteen yksityiskohtaisemman tarkastelun, kun tutkijan on mahdollista tutustua erikseen jopa jokaiseen yksittäiseen aineistoesiintymään. Vaikka laadullisen ja määrällisen tutkimuksen tutkimuskysymykset ja tavoitteet ovatkin erilaisia, niillä on kuitenkin paljon yhteistä ja ne täydentävät usein toisiaan. Palaamme tähän artikkelin loppupuolella (ks. myös Johdannon luku 4 tk.).

Laadullisen aineistopohjaisen kielentutkimuksen taustalla on usein käyttöpohjainen kielikäsitys, jonka mukaan kielellinen merkitys ja kielen rakenne perustuvat kielen käyttöön autenttisissa viestintätilanteissa. Esimerkiksi kahden toisilleen lähimerkityksisen sanan, kuten suomen *asti* ja *saakka*, eroja voidaan pyrkiä selvittämään kontekstin perusteella (Päiviö 2007). Usein tämän ajatuksen kohdalla viitataan Firthin (1968, 179) kuuluisiin sanoihin ”you shall know the meaning of a word by the company it keeps”, mutta jo Frege (1884, 116) esitti teoksessaan *Die Grundlagen der Arithmetik* vastaavia ajatuksia: ”Wir stellten nur den Grundsatz auf, dass die Bedeutung eines Wortes nicht vereinzelt, sondern im Zusammenhange eines Satzes zu erklären sei” (’Sanan merkitystä ei pidä tarkastella yksin, vaan lauseyhteydessä’). Kontekstin tarkastelussa huomiota voidaan kiinnittää esimerkiksi tutkittavien sanojen kanssa samassa tekstiyhteydessä esiintyviin ilmauksiin ja rakenteisiin, ympäröivän tekstin muodostamaan kokonaisuuteen sekä

tekstinulkoisiin, maailmantietoon perustuviin seikkoihin (ks. esimerkki 1). Laadullisen aineistopohjaisen kielentutkimuksen menetelmät voivat kyllä soveltua myös sellaiseen sanastontutkimukseen, jossa sanojen käyttöä ei oteta huomioon, kuten myöhemmin käyttämästämme esimerkistä 2 nähdään. Tällöin tutkittavia sanoja tarkastellaan tutkimusaineistoksi määritellyn joukon jäseninä, ja tutkimuskysymykset ja tutkimuksen tavoitteet määritellään toisin kuin perustapauksissa, joissa käyttökonteksti on mukana.

Laadullinen aineistopohjainen kielentutkimus soveltuu erityisen hyvin monitahoisten ilmiöiden syvälliseen ja kokonaisvaltaiseen tarkasteluun. Se on omiaan myös aiemmin vähän tutkittujen ilmiöiden kartoituksessa, sillä se mahdollistaa tutkimusaineiston yksityiskohtaisen käsittelyn sekä luokittelun rakentamisen ja tutkimuskysymysten tarkentamisen sen perusteella, mitä tutkimusaineisto paljastaa. Esittelemme tässä artikkelissa laadullista aineistopohjaista kielentutkimusta tietokanta- tai taulukkotyöskentelyn näkökulmasta, sillä kokemuksemme mukaan tällainen työskentelytapa helpottaa tutkimusaineiston käsittelyä ja lisää analyysin systemaattisuutta. Ennen laadullisen tutkimuksen prosessien kuvaukseen siirtymistä avaamme kuitenkin vielä lisää laadullisen aineistopohjaisen kielentutkimuksen historiallista taustaa.

**Mihin?**

Laadullinen aineistopohjainen kielentutkimus soveltuu kaikenlaisten (kooltaan kohtuullisten) tutkimusaineistojen tarkasteluun. Sen vahvuutena on taipuminen monitahoisten ilmiöiden tutkimiseen silloin, kun ilmiöstä halutaan saada syvällistä ja kokonaisvaltaista tietoa. Tietokanta- tai taulukkotyöskentely helpottaa tutkimusaineiston käsittelyä ja lisää analyysin systemaattisuutta.

## 2. Historiallista taustaa

Kielen kuvaaminen perustuu aina jollakin tavoin autenttiseen aineistoon. Abstrakteimmillaan ”aineisto” on yhtä kuin (äidinkielisen) tutkijan taju kielestä (ks. Kielitaju kielentutkijan työkaluna tk.), mutta tavallisempaa on, että aineisto on kerätty ja talletettu jossakin muodossa. Aineistojen muodot ovat kehittyneet käsin tehdyistä muistiinpanoista ja keräelmistä ääni- ja kuvatallenteisiin. Aiemmin fyysisessä muodossa tallennettuja kokoelmia on digitoitu ja digitoidaan edelleen aineistojen saatavuuden parantamiseksi. Laadullisen aineistopohjaisen kielentutkimuksen taustalla on ajatus siitä, että autenttisessa tilanteessa käytetty kieli antaa parhaat edellytykset tutkia valittua ilmiötä. Tällaisella aineistolla ajatellaan siis saatavan mahdollisimman hyvin todellisuutta vastaava kuva, jota aineiston tuottaja ei ole tietoisesti tai tahattomasti muovannut tutkimuksen tuloksiin vaikuttavalla tavalla. Kärjistetysti tämän ajatuksen voi ilmaista sanomalla, että aineisto tarjoaa autenttisen kuvan tutkijan tarkasteltavaksi, eriteltäväksi ja tulkittavaksi.

Suomen kielen tutkimuksessa laadullisella aineistopohjaisella tutkimuksella on pitkät perinteet, ovathan tieteenalan juuret olennaisella tavalla 1800-luvulla virinneessä kiinnostuksessa kansanrunouteen ja erityisesti itäisten murteiden edustamaan ”aitoon” ja ”turmeltumattomaan” kieleen[2] (esim. Häkkinen 1994, 441–442; Hovdhaugen, Karlsson, Henriksen & Sigurd 2000, 245–246; Häkkinen 2008, 89–91). Tuolloin aineiston kerääminen oli välttämätöntä, koska kokonaisvaltaista tietoa esimerkiksi murteiden ominaislaadusta ei yksinkertaisesti ollut (Häkkinen 2008, 101); aineistoja kerättiin kielen dokumentoimiseksi. Myöhemmin sanasto-, nimistö- ja kielennäytekokoelmia on kerätty eri yliopistojen suomen kielen tutkijoiden ja opiskelijoiden voimin, ja kenttätyö (esimerkiksi nimestys eli paikannimien keruu) kuului pitkään suomen kielen yliopisto-opintojen pakolliseen sisältöön (esim. Lyytikäinen 1984; Lyytikäinen & Yli-Paavola 2010; Suutari 2010). Kenttätyötä tehdään edelleen, mutta nykyisin työ on useimmiten johonkin tutkimushankkeeseen sitoutunutta (2000-luvulla esimerkiksi Kotimaisten kielten keskuksen murteenseuruuhanke[3] ja Turun yliopiston

Suomenkielisten arkikeskustelujen morfosyntaktinen tietokanta Arkisyn[4]; ks. myös Karttunen & Rouhikoski 2010).

Pitkään oli tyypillistä, että tutkijat keräsivät omat aineistonsa, joihin tutkimuksen saattoi perustaa (Häkkinen 2008, esim. 98–108, 117, 145). Kun aineistojen keruuta alettiin toteuttaa organisoidummin, syntyi myös kollektiiviseen käyttöön tarkoitettuja aineistoja, joita opiskelijat ja tutkijat voivat käyttää opinnäytetöihin ja muihin tutkimuksiin (korpuksista ks. Korpusaineistot tk.). Käsityksen aineistopohjaisen kielentutkimuksen aihepiirien kirjosta voi saada esimerkiksi Turun yliopiston suomen kielen ja suomalais-ugrilaisen kielentutkimuksen oppiaineista vuoden 1947 jälkeen valmistuneiden tutkielmien pro gradu -tietokannasta hakemalla vaikkapa aihepiirit ”sanastontutkimus” ja ”murretutkimus”[5] (ks. myös Leino, Majapuro & Kristiansson-Seppälä 1984; Lyytikäinen 1984, 450).

Suomen kielen tutkimuksen synty yhdistyi siis kielen dokumentointiin, ja tällä voi ajatella olleen vaikutusta siihen, miten keskeiseksi laadullinen aineistopohjainen tutkimus fennistiikan alalla muodostui. Perinteistä fennististä kielentutkimuksen tavoitetta onkin kuvattu luonnehdinnalla ”laaja, hyvin järjestetty aineisto puhuu puolestaan” (Ravila 1967, 35; ks. myös Karlsson 1975). Luonnehdinta heijastaa tutkimuksessa vallinnutta aineistouskoa ja arvostusta, jota aineiston parissa työskentelyä kohtaan tunnettiin. Näkemykseen saattoi liittyä myös käsitys siitä, että tutkimuksessa vastaan tulevat ongelmat voidaan poistaa hankkimalla ja käsittelemällä lisää aineistoa. Myöhemmin luonnehdintaan on suhtauduttu myös väheksyvästi: pelkkä aineiston kuvaaminen saati sen laajuus ei itsessään vie tutkimusta eteenpäin, vaan kiinnostavaa on vasta se, mitä aineistosta voidaan päätellä. Laadullisen aineistopohjaisen kielentutkimuksen menetelmän kuvaus voi kuitenkin hyödyntää perinteistä luonnehdintaa, ja tämä on lähtökohtamme seuraavassa jaksossa, jossa tarkastellaan menetelmää konkreettisemmin. On syytä muistaa, että tutkijan aikaisemmat tiedot (kielitaju, lingvistinen koulutus jne.) ovat aina jossakin määrin hänen tekemiensä valintojen taustalla: tutkimusaiheen, tutkimuskysymysten ja aineiston valinnassa, tutkimusaineiston koostumuksen arvioinnissa, esiintymien analyysissa ja päätelmien tekemisessä (ks. Kielitaju kielentutkijan työkaluna; Johdannon luku 4 tk.). Seuraava menetelmän kuvaus keskittyy kuitenkin siihen, miten toimia aineiston kanssa laadullisen kielentutkimuksen eri vaiheissa.

**Mitä tutkitaan?**

- kieliopin tutkimus
- sanaston ja terminologian tutkimus
- kielen oppimisen tutkimus
- kielihistoriallinen tutkimus jne.

# 3. Miten laadullista aineistopohjaista kielentutkimusta tehdään?

## 3.1. ”Laaja”: tutkimusaineiston hankkiminen ja koko

Ensimmäiseksi laadullista aineistopohjaista kielentutkimusta tekevän tutkijan on ratkaistava kysymys käytettävän tutkimusaineiston koostumuksesta ja laajuudesta. Tähän ei tietenkään ole yksiselitteistä vastausta, vaan vastaus riippuu tutkimuksen laajuudesta, tutkittavasta aiheesta ja myös saatavana olevista aineistoista. Jos haluaa tutkia jotakin ilmiötä vanhassa kirjakielessä, aineiston valinta saattaa olla hyvinkin suoraviivaista, jos aineistoja on niukasti ja niiden väliset genre-erot ovat pieniä. Näin on esimerkiksi vanhimmissa suomenkielisissä teksteissä, jotka kaikki edustavat uskonnollista käännöskirjallisuutta. Suurissa eurooppalaislähtöisissä kielissä tilanne voi tältäkin osin olla erilainen, koska kirjoitetun kielen historia on pidempi ja saatavana on eri tekstilajeja edustavia aineistoja (ks. Historiallinen korpuslingvistiikka tk.).

Nykykielisen ilmiön tutkijan sen sijaan on tehtävä valintoja esimerkiksi yleiskielisten aineistojen ja murre- tai keskusteluaineistojen välillä sekä kirjoitetun ja puhutun kielen aineistojen välillä. Kirjoitetun yleiskielen ja puhutun murteen tai arkikeskustelun välillä on luonnollinen yhteys, mutta saatavana on myös puhuttua yleiskieltä edustavaa aineistoa (esimerkiksi radiopuhe, esitelmä- ja väitöstilaisuudet, puheet) ja kirjoitettua murretta (murteella kirjoitettu kirjallisuus), ja tutkimus voi kohdistua myös siihen, miten keskustelua kirjoitetaan (sähköinen

vuorovaikutus, kaunokirjallisuuden dialogit, haastattelujen käyttö kirjoitetussa mediassa jne.). Tutkijan on myös päätettävä, haluaako hän vertailla ilmiön esiintymistä erilaisissa aineistoissa: miltä kirjoitetussa nyky-yleiskielessä esiintyvä ilmiö näyttää vaikkapa murteissa tai varhaisemmassa kirjakielessä? Useamman aineistotyypin käyttäminen ei automaattisesti tarkoita tutkimusaineiston laajuuden kasvattamista vaan ennemminkin sitä, että tutkimusaineiston koostumus harkitaan tarkasti.

Käytettävän tutkimusaineiston laajuuteen vaikuttaa myös se, millainen tutkittava aihe on. Onko ilmiö kielessä yleinen vai harvinainen? Onko tutkittavana yksi, yhtenäinen rakennetyyppi vai joukko toisiinsa jollakin tavoin kiinnittyviä rakenteita, esimerkiksi yksi postpositio vai postpositioiden ryhmä? Harvinaisen ilmiön tai yksittäisen rakenteen esiintymät saattaa olla mahdollista poimia tutkittavaksi kattavasti, mutta yleisestä ilmiöstä tai rakenteiden joukosta on poimittava otos, jolloin pitää valita, kuinka otos koostetaan (ks. Korpusaineistot tk.). Toisaalta ilmiö voi olla myös niin harvinainen, että sen esiintymiä ei löydä suuristakaan aineistoista, jolloin paras vaihtoehto tutkimusmenetelmäksi on introspektio (ks. Kielitaju kielentutkijan työkaluna tk.) tai kyselymenetelmän käyttö (ks. Kysely tk.).[6]

Käytettävän tutkimusaineiston kokoon vaikuttaa sekin, onko tutkimuksen kohteena oleva ilmiö poimittavissa jostakin sähköisestä aineistosta suoran aineistohaun avulla. Jos aineisto on morfologisesti ja syntaktisesti koodattu (kuten esimerkiksi Lauseopin arkiston[7] materiaali on), on haku mahdollista kohdistaa varsin tarkastikin määriteltyyn kieliopilliseen kategoriaan. Tällaisessa tilanteessa on hyödyllistä tutustua aineiston koodaamisessa noudatettuihin periaatteisiin, koska tutkijan tulkinta ilmiöstä voi olla erilainen kuin aineiston koostajien (esim. Ikola 1985 Lauseopin arkistosta; Inaba 2007 Mikael Agricolan teosten morfosyntaktisesta tietokannasta). Jos tutkimuksen kohteena on tietty sana tai tietyt sanat, ne on mahdollista hakea ainakin yleiskielisistä sähköisistä aineistoista merkkijonohaun avulla. Esimerkiksi Lauseopin arkiston aineisto on tätä nykyä myös hakusanoitettu, eli tietyn sanan esiintymät voi hakea yleiskielisen hakusanan avulla riippumatta siitä, mikä sanan äänteellinen asu kulloinkin on (esim. hakusanalla *tehdä* saa esiin mm. sanat *tehrä* ja *tehä*). Vanhempia kielimuotoja ja muita kuin yleiskieltä

edustavista (hakusanoittamattomista) aineistoistakin merkkijonohakuja voi tehdä, mutta silloin tulee huomata, että sanojen kirjoitusasut voivat olla moninaiset. Toisaalta vaikkapa metaforien aineistopohjainen tutkimus voidaan toteuttaa menetelmiä yhdistämällä: tutkija hankkii ensin käsityksen olemassa olevista metaforista tekstejä systemaattisesti läpi lukemalla, ja tämän vaiheen perusteella koostetaan sanajoukko, jonka esiintymät voidaan hakea korpuksesta automaattisesti (Tissari 2017). Hyvä esimerkki yksilöllisen tutkimusaineiston kokoamisesta ovat ns. imitatiivit, joiden hakeminen aitoa kielenkäyttöä edustavasta materiaalista vaatii tutkijan omaa tarkkaavaisuutta ja käsityötä (Jääskeläinen 2013). Myös vanhan kirjakielen tutkimuksessa tekstien läpi lukemista pidetään usein tärkeänä tutkimusaineiston keruumenetelmänä (koodattujen) digitaalisten aineistojen olemassaolosta huolimatta (esim. Nummila 2012).

Hakipa tutkimusaineistonsa millaisesta lähteestä ja millaisella keinolla tahansa, on poiminnat aina käytävä huolellisesti läpi sen varmistamiseksi, että ne todella kuuluvat tutkimuksen alaan. On tyypillistä, että hakutuloksen joukossa on roskaa, joka on siivottava käsin. Tutkimusaineiston käsittelyä ei voi jatkaa, ennen kuin siivous on tehty.

Laadullisessa aineistopohjaisessa kielentutkimuksessa analyysi tehdään pääosin käsityönä, ja tämän takia tutkimusaineiston koko on pidettävä kohtuullisena. On mahdotonta antaa yksiselitteisiä ohjeita siitä, minkä kokoinen aineisto on riittävä, koska se riippuu sekä tutkimusaiheesta että tutkimuksen laajuudesta. Tutkimuksen kuluessa voi luonnollisesti tulla vastaan tilanteita, joissa koostetun tutkimusaineiston supistaminen tai laajentaminen on paikallaan. Aineistoesiintymiä analysoidessaan voi jossakin vaiheessa havaita, että niukempi esiintymäjoukko riittää tutkimuskysymyksiin vastaamiseen. Toisaalta jos esimerkiksi selvitetään aiemmin kuvaamattoman kielen rakennetta, on aivan mahdollista, että valittu tai käytettävissä oleva tutkimusaineisto ei sisälläkään riittävästi tietoa jostakin ilmiöstä, vaan materiaalia on kuvan täydentämiseksi hankittava lisää (ks. Kielten dokumentointi ja kieliopin kuvaus tk.).

## 3.2. ”Hyvin järjestetty”: tutkimusaineiston käsittely

Aineiston järjestäminen tarkoittaa tutkimusaineistoon poimittujen esiintymien luokittelua ja koodaamista esiintymien ominaisuuksien perusteella. Koodaamisella[8] tarkoitamme tässä sitä, että tutkimusaineiston esiintymälle annetaan joukko sen ominaisuuksista kertovia nimilappuja sen mukaan, mitkä asiat esiintymässä kiinnostavat tutkijaa. Kaikkien menetelmien käytössä systemaattisuus on alku ja juuri, myös laadullisessa aineistopohjaisessa tutkimuksessa. On päätettävä, millaisia seikkoja esiintymistä koodataan ja miten, ja toteutettava koodaaminen järjestelmällisesti jokaisen esiintymän kohdalla. Tyypillisesti tutkittavaksi poimittuihin esiintymiin liitetään ainakin erilaisia metatietoja esiintymän alkuperäisestä lähteestä (esimerkiksi sanomalehtiaineiston tapauksessa ainakin julkaisupäivä, otsikko, lehden nimi, osasto jne.). Lisäksi voidaan koodata tutkimuskohteesta riippuen erilaisia lingvistisiä tietoja, kuten morfologian, syntaksin, semantiikan, pragmatiikan tai diskurssin tason merkintöjä. Nämä tiedot ovat kielikohtaisia, ja niiden koodaaminen edellyttää tutkittavan kielen rakenteen tuntemusta ja asianmukaisten lähteiden, esimerkiksi kielioppien, hyödyntämistä. Toimivan yksittäisestä kielestä riippumattoman yleiskuvauksen kielellisen rakenteen analyysista tarjoavat Payne (1997) ja Dixon (2010).

Esiintymän morfologisten ominaisuuksien analyysissa otetaan huomioon tutkittavan kielen taivutuskategoriat, kuten nomineilla luku ja sija. Sellaisissa kielissä, joissa on tämän tyyppistä taivutusta, lauseen syntaktinen analyysi on riippuvainen sanojen morfologisista ominaisuuksista, koska lauseenjäsenyys määräytyy morfologisen muodon perusteella (esimerkiksi suomessa objektin sijat ovat nominatiivi, partitiivi, yksikön genetiivi ja persoonapronomineilla akkusatiivi eikä muussa taivutusmuodossa oleva sana voi olla objekti). Lauseen syntaktisessa analyysissa on morfologisen muodon lisäksi katsottava lauseen kokonaisuutta ja sanojen suhteita toisiinsa. Lauseessa voi nimittäin olla useampia samassa taivutusmuodossa olevia sanoja, jolloin kokonaisuus ratkaisee, minä lauseenjäsenenä sana toimii. Suomen lauseessa *Pallo on pyöreä* molemmat nominit ovat yksikön nominatiivissa, mutta kokonaisuuden perusteella on selvää, että esineeseen viittaava substantiivi *pallo* toimii subjektina ja

ominaisuutta ilmaiseva adjektiivi *pyöreä* predikatiivina. Sanaluokka onkin perustana sanatason semanttiselle analyysille, jota voidaan tarkentaa tutkimuskohteen mukaan (vaikkapa jakamalla substantiivien ilmaisemat oliot edelleen elollisiin ja elottomiin tai ihmisiin, muihin elollisiin, esineisiin, rakennuksiin tms.). Lausetason semanttisessa analyysissa voi pohtia niin sanottuja semanttisia rooleja: esimerkiksi lauseen subjekti on tyypillisesti toiminnallinen agentti, objekti toiminnan kokeva kohde (patientti) ja adverbiaali jollakin tavoin toimintaa kehystävä taustaelementti, kuten sijainti (lokaatio; *Poni* [agentti] *veti kärryt* [patientti] *pihalle* [lokaatio]).

Tarkastellaan seuraavaksi yhtä tapaa koodata lauseen *Haahka on musta vatsan alta* morfologisia, syntaktisia ja semanttisia ominaisuuksia.

| | *Haahka* | *on* | *musta* | *vatsan* | *alta.* |
|---|---|---|---|---|---|
| sanaluokka | substantiivi | verbi | adjektiivi | substantiivi | postpositio |
| morfologia | yks. nom. | akt. ind. pr. y. 3. p. | yks. nom. | yks. gen. | yks. abl. |
| syntaksi | subjekti | predikaatti | predikatiivi | adverbiaali | |
| semantiikka | elollinen | *olla*-verbi/ kopula | väri | elollinen/ osa | osa/ kokonaisuus |

Harvassa tutkimuksessa tarvitaan näin perusteellista koodaamista tutkimusaineiston esiintymistä. Tärkeä osa tutkimusaineiston käsittelyä onkin koodaamisen pohtiminen suhteessa tutkimuskysymyksiin. Mitkä asiat ilmiössä kiinnostavat? Millaisten asioiden pohjalta niitä olisi mahdollista selvittää? Tutkimusaineiston käsittelyperiaatteiden huolellinen pohtiminen voi viedä eteenpäin paitsi konkreettista aineistotyötä myös tutkijan käsitystä tutkittavasta aiheesta ja tutkimuksen rakenteesta. Se, miten tutkimusaineistoa käsittelee, voi heijastua varsin suoraan siinä, miten tutkimuksesta raportoi tekstinä. Se voi myös auttaa ymmärtämään tutkimuksen teoriasta kumpuavia käsitteitä, kun pohtii, mitkä niistä olisivat tutkimusaineiston koodaamiseen soveltuvia (ks. muuttujan ja kiintopisteen koodaaminen esimerkissä 1). Näin teoria myös suuntaa analyysia tarjoamalla käsitteitä tutkittavan ilmiön analysointiin. On tutkijan valinta, kuinka vahvaa käytetyn teoriataustan ohjaus

aineiston käsittelyssä on. Täysin teoriasta vapaata se ei ole koskaan.

Laadullisen aineistopohjaisen kielentutkimuksen keskeinen taustaoletus on, että aitoon kielenkäyttöön perustuva aineisto on arvokasta ja jopa välttämätöntä kielentutkimukselle. Tästä seuraa myös se, että tutkittavat ilmiöt ovat sidoksissa toisiinsa. Kielentutkimuksen tärkeimpiä tavoitteita onkin selvittää, millaisia ilmiöiden (esimerkiksi morfosyntaktisten ominaisuuksien) väliset suhteet ovat: jos tarkastellaan ilmiötä X (esimerkiksi jotakin postpositiota), mitä säännöllisiä seuralaisia sillä on (esimerkiksi millaisten verbien kanssa ko. postpositio viihtyy)? Myöhemmässä vaiheessa päästään myös pohtimaan, miksi eri elementit viihtyvät yhdessä ja mitä tällaisesta kumppanuudesta seuraa. Tutkimusaineiston käsittelyssä onkin aina kiinnitettävä jossain määrin huomiota myös ilmiön esiintymisympäristöön ja koodattava jonkin verran myös kielellisen kontekstin ominaisuuksia.

Edeltävä esimerkki on mukaelma esiintymästä Ojutkankaan (esim. 2008a) käyttämässä tutkimusaineistossa, jonka avulla tutkitaan suomen paikkaa ilmaisevien postpositioiden ja adverbien käyttöä ja merkitystä. Tutkimusaineistoon kuuluu murreaineistoa (Lauseopin arkistosta), vanhaa ja 1800-luvun kirjakieltä (Kotimaisten kielten keskuksen vanhan kirjasuomen ja varhaisnykysuomen korpuksista) ja nyky-yleiskieltä (mm. Kielipankista). Tutkimuksessa koodaus toteutuu seuraavaan tapaan:

| | *Haahka* | *on* | *musta* | *vatsan* | *alta.* |
|---|---|---|---|---|---|
| sanaluokka: | | | | substantiivi | postpositio |
| morfologia: | | | | yks. gen. | yks. abl. |
| syntaksi: | subjekti | | | | |
| semantiikka: | elollinen | *olla*-verbi | | elollinen/ osa | osa/ kokonaisuus |
| | muuttuja | | | kiintopiste | |

Tutkimusaineiston koodaamisessa on suositeltavaa käyttää jotakin tietokanta- tai taulukkolaskentaohjelmaa. Siitä on etua monessa suhteessa.

Ensinnäkin tietokanta- tai taulukkomuoto auttaa koodaamisen systemaattisuudessa, kun jokaisella tutkimusaineiston esiintymällä on oma paikkansa, jossa se odottaa nimilappujensa täyttämistä. Toiseksi sähköisessä muodossa oleva tutkimusaineisto pysyy tallessa ja järjestyksessä, on kopioitavissa ja kulkee kevyesti paikasta toiseen tarpeen mukaan, kaikkein kevyimmin pilveen tallennettuna. Siihen on myös helppo tehdä hakuja, mikä auttaa löytämään tarkastelun kohteena olevat tapaukset nopeasti. Kolmanneksi tietokanta- tai taulukkomuoto voi auttaa merkittävästi ilmiön ominaisuuksien välisten suhteiden hahmottamisessa, toisin sanoen tutkimustulosten saavuttamisessa (tähän palaamme seuraavassa jaksossa). Käytännössä sähköinen muoto rakennetaan siten, että kukin aineistoesiintymä sijoitetaan omalle rivilleen ja esiintymän lähdetiedoille ja kullekin esiintymästä analysoitavalla seikalle varataan taulukosta oma sarake. Sarakkeet täytetään systemaattisesti tutkijan valitsemilla yksiselitteisillä koodeilla. Oheinen esimerkki havainnollistaa edellä käytettyä esimerkkiä hyödyntäen, millaisia tietoja tietokanta tai taulukko voi sisältää.

Esimerkki 1 koodauksesta:

| **Numero** | **Grammi** | **Esimerkki** | **Lähde** | **Käsite** | **Käsite 2** | **Sija-muoto** |
|---|---|---|---|---|---|---|
| 4 | AL | VL Niin koiranen- o/ **valkone *seljäm <<päält>> ja/ ja **musta/ *vatan- <<alt>> ja **naaranen- o *semmone/ *ruskia**harma.// | LA/murre LUVIA 0011 227 | paikka | osa-kokonaisuus | E-abl |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |

| **Syntaktinen rooli** | **Verbi** | **Muuttuja** | **KP: morfologia** | **KP: semanttinen luokka** | **Sanajärjestys** | **Muut grammit** | **Huomautuksia** |
|---|---|---|---|---|---|---|---|
| postpos | olla | elollinen, sb | leks-gen | elollinen/osa | | päältä | haahka |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |

1) Esiintymän juokseva numero tulee suoraan tietokantaohjelmasta. Se on muuttumaton tieto, joka varmistaa, että tietue (eli yhden rivin tiedot) pysyy koossa.

2) Tietokannassa on useiden eri grammien eli adpositioitten (pre- ja postpositioitten) ja adverbien esiintymiä (*grammi* on kätevä suomen kieliopin tutkimukseen vakiintunut kattonimitys sanoille, jotka toimivat sekä adpositioina että adverbeina; termin on alkuaan ottanut käyttöön Svorou [1993]). Suomen grammeilla on useimmiten vähintään kolme taivutusmuotoa, ja tähän sarakkeeseen kirjatun koodin avulla on helppo hakea kaikki grammin muodot (tässä tapauksessa *alla*, *alta* ja *alle* eri murrevariantteineen).

3) Grammin esiintymävirke (tarvittaessa laajemmankin kontekstin kera). Tarkastelun kohteena oleva elementti nousee tekstistä esiin kaksinkertaisten kulmasulkeiden sisällä. Jos esiintymävirkkeessä on useita grammeja, virke toistuu tietokannassa. Tämä esiintymä siis on tietokannassa kahdesti, AL- ja PÄÄ-koodilla. Esiintymän alussa ovat puhujan nimikirjaimet erottamassa haastattelijan ja haastateltavan puhetta toisistaan.

4) Esiintymän lähdetiedot. Tässä LA tarkoittaa Lauseopin arkistoa ja Luvia pitäjää. 0011 on Luvian koodi Lauseopin arkistossa ja 227 virkkeen numero Luvian näytteen tekstissä. Lähdetietojen perusteella voisi tässä tehdä päätelmiä siitä, esiintyykö grammin jokin käyttötapa vain tietyllä murrealueella. Jos aineistona olisi eri tekstilajeja, lähdetiedon perusteella voisi tehdä vastaavasti havaintoja tekstilajien välisistä eroista tarkasteltavan ilmiön osalta.

5 ja 6) Esiintymän ensisijainen merkitys ja mahdollinen merkityksen tarkennus. Tämän avulla on mahdollista tehdä esiintymien siivouksen loppusilaus, esimerkiksi karsia pois tapaukset, joissa grammin merkitys ei olekaan sijaintia ilmaiseva. Näiden tietojen perusteella saa luonnollisesti esiin myös merkitysryhmiä osana tutkimustuloksia.

7) Grammin morfologinen muoto. Tässä E tarkoittaa erosijaa; tämän koodin avulla on helppo poimia kootusti elatiivi-, ablatiivi- ja partitiivimuotoiset grammit, jos tarpeen. Abl tarkoittaa ablatiivia. (Näiden tietojen käytöstä tutkimuksessa on esimerkki luvussa 3.3.)

8) Grammin syntaktinen rooli postpositiona (yhdessä genetiivisijaisen täydennyksen kanssa) tai adverbina (yhdessä paikallissijaisen seuralaisen kanssa tai ilman seuralaista). Grammin syntaktinen rooli kytkeytyy suoraan kiintopisteen ilmaisemiseen (ks. kohta 11).

9) Verbin semanttinen luokka. Koodaus on tehty Anneli Pajusen suomen kielen verbejä käsittelevän *Argumenttirakenne*-teoksen (2001) luokitusta soveltaen. Kun tutkii sanojen käyttöä, verbin tyyppi (esim. liikeverbi, antamisverbi) on usein äärettömän hyödyllinen tieto. Se auttaa usein ratkaisevasti hahmottamaan esiintymien merkitystä ja esiintymille mahdollisia konstruktioita. Jos tutkittava aineisto on lausemuodossa, verbi kannattaa koodata aina.

10) *Muuttuja* on kognitiivisen kielentutkimuksen käsite, joka tarkoittaa huomion ensisijaisena kohteena olevaa osallistujaa. Merkityksen analyysissa se on siis varsin keskeinen koodattava. Tähän sarakkeeseen koodataan grammin muuttujan semanttinen luonnehdinta. Tässä tapauksessa on kyse koirashaahkan värityksestä, joten koodaus on elollinen. Samaan kenttään on koodattu myös muuttujan syntaktinen asema (sb = subjekti), mikä ei ole aivan viisasta. Aina olisi hyvä, että jokaisella tiedolla olisi oma sarake.

11) *Kiintopiste* on kognitiivisen kielentutkimuksen käsite, joka tarkoittaa huomion toissijaisena kohteena olevaa osallistujaa. Grammien tutkimuksessa kiintopisteen analyysi on melkeinpä muuttujaakin keskeisempi, koska se on grammien esiintymäkonstruktioissa ratkaisevan tärkeä jäsen. Tähän koodataan kiintopisteen ilmaisukeinon muoto. Leks tarkoittaa substantiivia, joka on yleisnimi, ja gen genetiiviä.

12) Kiintopisteen semanttinen luokka. Vatsa on koodattu elollisen osaksi. Ruumiinosa olisi myös tähän mahdollinen koodaus, mutta tällä

tavoin erotetaan ihmisen ja muiden eläinten osat toisistaan. Ihminen on kuitenkin monella tavoin erityinen puheenaihe, ja siksi on hyödyllistä pitää se koodauksessa omana ryhmänään. Kiintopisteen semanttinen luokka on analyysissa ratkaisevan tärkeä seikka grammin merkityksiä ja käyttöalaa tarkasteltaessa.

13) Kun tutkii sanojen käyttöä konstruktioissaan, voi sanajärjestys toisinaan olla tarpeellinen tieto. Juuri tämän esiintymän kohdalla mitään kommentoitavaa ei toistaiseksi ole ollut.

14) Tutkimuskohteesta riippuen voi olla hyödyllistä koodata erikseen näkyviin jokin seikka esiintymäympäristöstä. Tässä tapauksessa samassa yhteydessä esiintyvien grammien analyysi on osoittautunut kiinnostavaksi.

15) Huomautuksia-sarakkeeseen voi kirjata erilaista täydentävää tietoa ja koodauksen kuluessa herääviä kysymyksiä ja huomioita. Tähän on kirjattu esiintymän kontekstista varmistettu tieto, että puheen aiheena on haahkojenväritys.

Yllä kuvatussa jäsennyksessä on kyse syntaktis-semanttisesta tutkimuksesta, ja siksi koodattavia seikkoja on runsaasti. Tämä ei tietenkään aina ole tarpeen. Havainnollistamme seuraavalla esimerkillä 2 sanasemanttisessa tutkimuksessa käytetyn aineiston jäsennystä ja koodausta (Ojutkangas 2010).

Esimerkki 2 koodauksesta:

| **Tunniste** | **Eläin-eläin** | **Sanan tyyppi** | **Luokka** | **Lahko** | **Määrite-osa** | **Perusosa** | **M-osan luokka** |
|---|---|---|---|---|---|---|---|
| 160 | orava-käki | virall. lajin-nimi | lintu | käet | orava | käki | nisäkäs |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

| M-osan motivaatio | Tieteellinen nimi | Englanniksi | Huomioita |
|---|---|---|---|
| toiminta: juoksee pitkin oksia ja hyppää oksalta toiselle oravien tapaan, lentää vain lyhyitä matkoja (http://en.wikipedia.org/wiki/Squirrel_Cuckoo 21.8.2009) | Piaya cayana | Squirrel Cuckoo | |
| 9 | 10 | 11 | 12 |

1) Kuten esimerkissä 1.

2) Tutkimuksen kohteena oleva sana ilman kontekstia, koska sanojen käyttö ei ollut tutkimuksen kohteena. Sanat kerättiin saatavana olevista virallisen nimistön listoista (esim. Väisänen, R. A., Högmander, H., Björklund, H., Hänninen, L., Lammin-Soila, M., Lokki, J., Rauste, V. 2006: *Maailman lintujen suomenkieliset nimet*. (2. uudistettu p.); BirdLife Suomi: https://www.birdlife.fi/lintutieto/maailman-lintulajien-suomenkieliset-nimet/.)

3) Sanatyypillä seulottiin tutkimuskohteeksi nimenomaan lajinnimet ja jätettiin esimerkiksi heimojen nimet (kuten *lehmäantiloopit*) tutkimuksen ulkopuolelle.

4) Tähän koodattiin lajin biologinen luokka, joka toimi tutkimusaineiston perusryhmittelynä.

5) Lahko on luokkaa tarkempi biologinen ryhmittely, joka ei tässä tutkimuksessa lopulta ollut tarpeen.

6) Tähän kirjattiin yhdyssanan määriteosana (eli ensiosana) oleva eläimen nimi, joka oli varsinaisen kiinnostuksen kohteena.

7) Yhdyssanan perusosan (eli viimeisen osan) kirjaaminen toimi parina edelliselle kohdalle mutta ei lopulta ollut kovin tarpeellinen erittely.

8) M-osan luokka tarkoittaa määriteosan nimeämän eläimen biologista luokkaa. Tärkeä pari kohdalle 4.

9) Tähän kirjattiin määriteosan motivaatio: mitä määriteosa kertoo lajista? Esimerkiksi *orava käen* määritteenä perustuu toimintaan, ennen kaikkea liikkumisen tapaan, ei esimerkiksi ravintoon (tämä laji ei syö oravia eikä oraville tyypillistä ravintoa) tai käen munien hautojaan ja poikasten kasvattajaan (oravakäet hautovat munansa ja hoitavat poikasensa itse). Näiden tietojen kokoamisessa käytettiin runsaasti eri lajeja käsittelevää lähdekirjallisuutta sekä internetin luotettavia biologisia lähteitä.

10) Lajeista puhuttaessa on tärkeää kuljettaa tutkimuksessa mukana lajien tieteellisiä nimiä, koska niiden avulla lajien jäljille pääsee, vaikka suomenkielinen nimi muuttuisikin.

11) Englanninkielinen nimi vihjaa, että lajinnimessä on kyse käännöslainasta, ja alun perin tavoitteena oli kirjata myös tieto lajin ruotsinkielisestä nimestä. Lajien nimeämisen tämän puolen tarkastelu jäi kuitenkin tutkimuksen ulkopuolelle.

12) Kuten esimerkissä 1.

Tutkimusaineiston koodaaminen käsin voi tuntua puuduttavalta ja mekaaniselta, mutta se on erittäin hyvä tapa tutustua kerättyihin esiintymiin ja syventyä tutkittavaan teemaan. Manuaalisen koodaamisen ehdoton etu on sen tarkkuus verrattuna automaattisiin annotointiohjelmiin ja parsereihin, joiden tarkkuus riippuu kielestä, tekstilajista ja annotoitavasta asiasta (annotointitarkkuuksista ks. Haverinen ja muut 2014; Dozat, Qi & Manning 2017). Toisaalta myös ihminen tekee koodaamisessa virheitä. Koodaamisessa onkin hyväksyttävä tietty epävarmuus ja se, että tarkinkaan koodaaja tuskin saa kerralla valmista. Ei siis ole mitään syytä tuskastua, jos johdonmukaisuus tuntuu aluksi vaikealta tai jos joutuu koodaamaan joitakin osioita tutkimusaineistostaan useaan kertaan. On päinvastoin luonnollista, että tutkimuksen edetessä herää kysymyksiä, jotka edellyttävät jonkin uuden asian selvittämistä, ja tällöin

voi joutua käymään koko tutkimusaineiston läpi uudelleen. Lisäksi etenkin monitulkintaisten ilmausten kohdalla tutkimusaineisto ei useinkaan asetu ongelmitta toisensa poissulkeviin luokkiin. Tällöin voi olla hyödyllistä sallia tilanne, jossa yksittäinen ilmaus voidaan luokitella kahden eri kategorian alle. (Esim. Vartiainen 2016, 45–46.)

Laadullisen aineistopohjaisen kielentutkimuksen menetelmän etu on, että se joustaa tutkimuksen etenemisen ja tutkijan uteliaisuuden mukana. Tällainen kehkeytyvä tutkimusasetelma (engl. *emergent research design*, ks. esim. Creswell 2014) mahdollistaa sen, että tutkija voi tarvittaessa muokata koodaustapaansa tai jopa tutkimusaineistonsa koostumusta tutkimuksen edetessä kertyneen tiedon perusteella. Näin tutkijan ei tarvitse sitoutua tiukasti tutkimuksen alussa tekemiinsä ratkaisuihin, mikäli myöhemmin osoittautuu, että ne eivät ole toimivia.

Entä onko kuhunkin sarakkeeseen täytettävät nimilaput eli koodit syytä päättää etukäteen? Vastaus tähänkin kysymykseen riippuu tutkimuksen kohteesta ja siinäkin vielä erikseen koodattavasta seikasta. On seikkoja, joissa etukäteen päätetyt kategoriat helpottavat koodaamista. Tällaisia ovat esimerkiksi morfologiset ja syntaktiset ominaisuudet. Semanttisten ominaisuuksien osalta sen sijaan voi olla hyödyllistä, että antaa tutkimusaineiston ohjata kategorioiden syntyä. Tällaisessa koodaustavassa annetaan siis ensin semanttisten luonnehdintojen syntyä samalla kun esiintymiä käydään läpi. Tällaisten tutkimusaineiston havainnointiin perustuvien luonnehdintojen perusteella voidaan sitten toteuttaa varsinainen koodaus, jossa ensikoodauksen ryhmiä voidaan yhdistellä yleisempien nimitysten alle tarpeen mukaan. Ensikoodauksen ei tarvitse kohdistua koko tutkimusaineistoon, vaan usein varsinaisten ryhmien muodostamiseen riittää, että käy läpi osan esiintymistä ja havainnoi ja koodaa niiden ominaisuuksia. Toisinaan voi myös olla hyödyllistä tehdä kattavampikin pilottitutkimus jollakin esiintymäjoukolla tutkimusaineiston jäsennyksen ja koodauksen toimivuuden testaamiseksi. Pilottitutkimuksen perusteella voi koodausta muokata ennen kuin ryhtyy tekemään varsinaista analyysia.

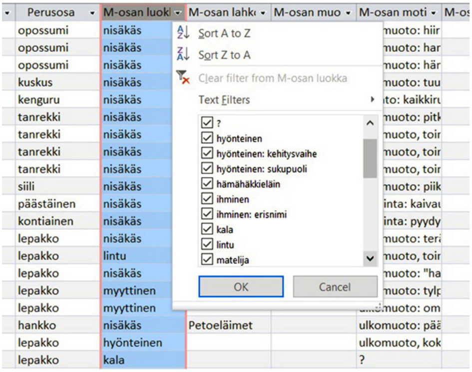
Pilottitutkimus voi olla erityisen tarpeellinen silloin, jos tutkimusaineiston järjestämisen apuna käytetään toista koodaajaa. Toisen koodaajan käyttäminen nopeuttaa tutkimusaineiston koodaamista ja voi

olla hyödyllistä myös koodauksen laadunvarmistuksessa. Toista koodaajaa käytettäessä tutkimusaineiston järjestämisperiaatteiden tulee olla erityisen selkeästi ja tarkasti kuvattuja, jotta aineiston koodaus säilyy yhtenäisenä.

## 3.3. "Puhuu puolestaan": tutkimusaineiston analysointi

Systemaattisen tutkimusaineiston parissa työskentelyn parhaita puolia on se, että työskentelyn kuluessa melkeinpä väistämättä nousee esiin vihjeitä analysoitavien asioiden välisistä suhteista. Juuri ilmiöiden välisten vuorovaikutussuhteiden etsiminen on keskeinen kielentutkimuksen tavoite. Laadullisessa aineistopohjaisessa kielentutkimuksessa tutkija paneutuu inhimillisesti järkevän kokoiseen materiaaliin, josta pyritään hankkimaan systemaattista tietoa koodauksen avulla. Tutkijan tehtävänä on tarkkailla, kuunnella ja tulkita tutkimusaineistoa mahdollisimman ennakkoluulottomasti ja välittää havaintonsa tutkimusraportissaan. Kielentutkijoiden saatavilla on kuitenkin yhä enemmän tutkimusaineiston käsittelyä helpottavia ja nopeuttavia automaattisia työkaluja (ks. tietolaatikko *Laadullisen analyysin tekniset apuneuvot*), joiden ansiosta yhä suurempien tutkimusaineistojen laadullinen analyysi tulee mahdolliseksi.

Tietokanta- tai taulukkolaskentaohjelmissa on mahdollista käyttää ns. suodattimia (engl. *filters*). Niiden avulla saa ensinnäkin näkyviin sen, millaisia koodauksia kussakin sarakkeessa on käytetty, ja voi varmistaa, että koodaus on johdonmukainen (kuva 1). Kuvassa 1 suodatin näyttää osan koodauksista, joita on käytetty edellä kuvatussa kahdesta eläimen nimestä koostuvien lajinnimien tietokannassa määriteosan luokan analysoimiseen. On hyvä huomata, että koodien joukossa on myös kysymysmerkki: tutkimusaineiston analyysiin jää väistämättä kohtia, joita on mahdoton selvittää ainakaan yksiselitteisesti. Tämä kuuluu kielen luonteeseen, sillä kielen kategoriat ovat harvoin jyrkkärajaisia ja kristallinkirkkaita.




Kuva 1. Kuvakaappaus tietokantaohjelman suodatintoiminnosta.

Suodatin on yksinkertaisempi versio tietokantaohjelmien sisältämistä kysely-toiminnoista. Kyselyssä on mahdollista määrittää samanaikaisesti useampaan sarakkeeseen kohdistuvat hakuehdot, joiden avulla voi tarkastella lähemmin jotakin aineiston osaa. Toki myös suodattimia on mahdollista käyttää kerroksittain, siten, että suodattaa useamman sarakkeen tietoja samanaikaisesti.

Eri sarakkeiden välisten koodausten yhteydet tuovat esiin tutkimusaineiston tarjoamia tuloksia. Jo koodatessa voi tehdä alustavia havaintoja mahdollisista suhteista, joiden olemassaolon haluaa tarkistaa, kun koodaus on valmis, mutta tutkimusaineistolla kannattaa myös ihan vain leikkiä: tehdä erilaisia hakuja, kyselyitä ja suodatuksia ja etsiä sillä tavoin toisiinsa vaikuttavia seikkoja. Jos tutkittava aineisto on lause- tai virkemuodossa, miltä vaikkapa eri verbiryhmien ympärille rakentuneet esiintymät näyttävät?

Esimerkiksi kahdesta eläimen nimestä koostuvissa lajinnimissä havaittiin yhteys tarkasteltavan lajin luokan ja määriteosan luokan välillä: kunkin luokan lajinnimissä painottuvat saman luokan jäseniin viittaavat nimet (siis nisäkkäiden nimissä nisäkkäät, lintujen nimissä linnut jne., esim. *vuohiantilooppi, allihaahka*). Toisin sanoen havainnolliset nimitykset otetaan pääosin saman eliöryhmän piiristä, kun lajeja nimetään toisen eläimen avulla. (Ojutkangas 2010.) Tämä esimerkki tuo esiin myös keskeisen menetelmien välisen yhteyden: laadullinenkin

analyysi hyödyntää usein jossakin määrin myös määrällisen analyysin menetelmiä (ks. Johdannon luvut 1 ja 4 tk.; Määrällinen korpuslingvistiikka tk.). Laadullisessakin tutkimuksessa määrälliset tiedot toimivat tyypillisesti tutkimuksen rajauksen perusteena tai yleisemmin tutkimuksen lähtökohtana, kun tutkimusaineistossa yleisiksi osoittautuneita ilmiöitä lähdetään tarkastelemaan lähemmin ja niille etsitään perusteita muualta kielen rakenteesta tai kielenulkoisesta maailmasta. Silloin kun tutkimusaineisto on kerätty systemaattisesti koostetusta aineistosta, on mahdollista myös suhteuttaa kielellisiä ilmiöitä niiden yleisyyden perusteella (onko ilmiö kielessä yleinen vai harvinainen) tai verrata ilmiön esiintymistä erilaisissa tekstityypeissä (yleiskieli vs. erikoiskieli, eri-ikäiset kielimuodot). Laadullisessa aineistopohjaisessa kielentutkimuksessa määrälliset tiedot ja varsinaiset tilastolliset menetelmät eivät kuitenkaan ole tutkimuksen keskiössä.

Kaikki laadullisen aineistopohjaisen kielentutkimuksen esiin tuomat seikat eivät välttämättä liity ilmiöiden välisiin suhteisiin, vaan muunkinlaista tietoa voi luonnollisesti nousta esiin. Esimerkiksi ensimmäisen käyttämämme esimerkin *Haahka on musta vatsan alta* pohjalta voi tehdä havainnon suomen ablatiivisijan merkityksestä: grammin muotona myös sillä voidaan ilmaista osa–kokonaisuus-suhdetta, joka on toisen erosijan, elatiivin, keskeisiä merkityksiä (vrt. *musta vatsasta* mutta ei *vatsalta*).

**Laadullisen analyysin tekniset apuneuvot**

Vaikka laadullista aineistopohjaista tutkimusta voi tehdä täysin manuaalisestikin, kielentutkijoiden saatavilla on lukuisia laadullista analyysia helpottavia teknisiä apuvälineitä. Aineiston keräämiseen, koodaamiseen ja tallentamiseen soveltuvia työkaluja ovat esimerkiksi erilaiset taulukkolaskenta- ja tietokantaohjelmat, kuten useimpiin Microsoft Office -paketteihin kuuluvat MS Excel ja MS Access tai vastaavat vapaasti ladattavissa olevat Apache OpenOffice Base- ja MySQL-ohjelmat. Laadullisen analyysin tarpeisiin on kehitetty myös omia kaupallisia ohjelmia, kuten NVivo. NVivon vahvuutena on se, että se soveltuu hyvin kokonaisten tekstien

ja kuvien analyysiin, mihin tietokanta- ja taulukkolaskentaohjelmat eivät yleensä sovellu yhtä hyvin.

Kielentutkijoille suunnattuja ohjelmistoja ovat esimerkiksi Laurence Anthonyn vapaasti ladattava AntConc-ohjelmisto sekä kaupallinen WordSmith Tools. Ohjelmistojen erilaiset työkalut auttavat löytämään kielessä usein toistuvia kuvioita, kuten kollokaatioita ja avainsanoja (ks. Määrällinen korpuslingvistiikka tk.). Tällainen tarkastelu voi olla hyödyllistä esimerkiksi tutkimuksen alkuvaiheen ideoinnin apuna, mutta nämä ohjelmistot tarjoavat myös määrällisiä tietoja, jotka voivat osoittautua merkityksellisiksi aineiston analysoinnissa (ks. Historiallinen korpuslingvistiikka tk.).

## 4. Tulosten tulkinta

Aitoja kielenkäyttöaineistoja hyödyntävissä tutkimuksissa tutkijan aineistolle antama rooli voi vaihdella. Kirjallisuudessa on käytetty kolmijakoa aineistoesimerkein tuettuihin (engl. *corpus-illustrated/corpus-exemplified*) (Tummers, Heylen & Geeraerts 2005; Glynn 2014), aineistopohjaisiin (engl. *corpus-based/corpus oriented*) ja aineistovetoisiin (engl. *corpus-driven*) tutkimuksiin (Tognini-Bonelli 2001; Glynn 2014) (ks. myös Jantunen 2009; Ojutkangas 2008b; Luodonpää-Manni, Penttilä & Viimaranta 2017). Aineistoesimerkein tuetuissa tutkimuksissa kieliaineistoa hyödynnetään pääasiallisesti tutkijan omaan kielitajuun perustuvan analyysin havainnollistamisessa. Tällainen tapa käyttää aineistoa pelkästään esimerkkien tarjoamisessa on kuitenkin suomalaisessa kielentutkimuksessa melko harvinainen. Aineistopohjaisessa tutkimuksessa tutkijan kielitajulla on myös oma roolinsa tutkimuksen suunnittelussa ja tutkimuksen lähtökohtien määrittelyssä (esimerkiksi tutkimusaiheen valinta ja tutkimuskysymyksen muotoilu voivat perustua tutkijan tekemään havaintoon kielestä tai aikaisempaan

tutkimukseen), mutta toisin kuin aineistoesimerkein tuetuissa tutkimuksissa aineistopohjaisissa tutkimuksissa korpusaineistoja analysoidaan ja havainnoidaan systemaattisesti. Aineistopohjainen tutkimus soveltuu hyvin suhteellisen laajojen kieliaineistojen analysointiin, ja se onkin hyvin suosittu tapa tehdä sekä laadullista että määrällistä tutkimusta. Aineistovetoiset tutkimukset eroavat aineistopohjaisista tutkimuksista mm. siten, että siinä missä aineistopohjaisissa tutkimuksissa analyysi voi perustua esimerkiksi tutkimuksen teoriasta kumpuaviin luokituksiin, aineistovetoisissa tutkimuksissa analyysikategorioiden annetaan muotoutua aineiston itsensä perusteella. Aineistovetoiset tutkimukset ovat yleisesti ottaen tavallisempia määrällisissä korpustutkimuksissa (ks. Määrällinen korpuslingvistiikka tk.), sillä tarkastelussa käytetään tyypillisesti tilastollisia menetelmiä, joilla pyritään selvittämään kielellisten ilmiöiden yleisyyttä kielenkäytössä.

Tässä artikkelissa on kuvattu aineistopohjaisen kielentutkimuksen menetelmää, mutta siinä on myös aineistovetoisen tutkimuksen piirteitä, sillä osa analysoitavista kategorioista muodostetaan aineistossa havaittujen piirteiden pohjalta. Tutkijan tuleekin itse ratkaista, missä määrin hän tukeutuu tutkimuksen teoriataustan ja aikaisempien tutkimusten tarjoamiin kategorioihin ja missä määrin hän antaa aineistosta tekemiensä havaintojen ohjata analyysia. Valintaan vaikuttaa mm. tutkimuskysymys: halutaanko esimerkiksi testata olemassa olevia teorioita vai etsiä aikaisemmin kartoittamattomia ilmiöitä?

Vaikka kovin tarkan rajan vetäminen erityyppisten tutkimusten välille voi olla turhaa, tutkijan on kuitenkin hyvä olla tietoinen siitä, millaisen roolin hän antaa tutkimusaineistolleen ja mikä hänen oma roolinsa tutkimuksessa on tiedonkäsittelijänä. Aineistopohjaisuus sen paremmin kuin aineistovetoisuuskaan eivät tarkoita teoriatonta tutkimusta: kategoriat eivät nouse aineistosta itsestään, vaan tarvitaan aina tutkijan suorittamaa tulkintaa tutkimusaineiston olennaisista piirteistä. On hyvä huomata, että tämä tulkinta ei synny tyhjästä, vaan tutkijan aikaisemmat tiedot tutkimuskohteesta, teoreettinen tausta ja metodologiset valinnat ohjaavat myös tutkittavan ilmiön havainnointia (ks. Niiniluoto 1980, 224–225). Tutkimuksen teoreettiset ja metodologiset ratkaisut tuleekin kirjoittaa huolellisesti auki, jotta niitä voidaan arvioida avoimesti.

Konkreettisesti tulosten tulkitseminen tarkoittaa tutkimusaineiston analyysivaiheessa tehtyjen havaintojen kirjoittamista tekstin muotoon. Kirjoittamalla konkretisoidaan ilmiöiden välisiä suhteita ja pohditaan, miten ne voidaan selittää. Koska kirjoittaminen on prosessi, ei kannata arkailla kirjoittaa näkyviin oudoilta, triviaaleilta tai epäuskottaviltakin tuntuvia seikkoja. Niiden rooli tulosten tulkinnassa varmistuu vasta työn edetessä. Usein voi olla niin, että kaikkein itsestäänselvimmältä tuntunut havainto on tulosten kannalta perustavanlaatuisin. On myös hyvä huomata, ettei tutkimusprosessi useinkaan ole täysin lineaarinen, niin että vaiheet seuraisivat toisiaan limittymättä. On tavallista, että analyysia ja tulkintaa tehdään vuorovaikutuksessa; jokin tulkinta voi esimerkiksi tuottaa tarpeen analysoida tutkimusaineistoa uudesta näkökulmasta, ja jonkin havainnon tekeminen voi tehdä kirjoittamisen tarpeelliseksi, vaikkei analyysi vielä olisi täysin valmis.

Usein on hyödyllistä kirjoittaa varsin yksityiskohtaisestikin näkyviin ratkaisut, joita on tehnyt tutkimusaineiston käsittelyssä. Miten esimerkiksi lopullinen tutkimusaineisto rajattiin? Tämä kytkeytyy tiiviisti tutkimuskysymysten tarkentumiseen. Myös käytettyjen suodattimien tai kyselyjen kuvaaminen tekstin muodossa voi auttaa tutkijaa hahmottamaan, miten tekstiä kannattaa rakentaa; tutkimusaineiston luokittelutapaa voi siis käyttää hyödyksi tutkimuksen rakennetta suunniteltaessa. Tekstin viimeistelyvaiheessa sitten tehdään päätös siitä, mikä osa tällaisesta tutkimusprosessin kuvauksesta on myös lopullisen työn lukijalle hyödyksi. Tutkimuksen mahdollisimman suuri läpinäkyvyys helpottaa tieteellistä laadunvarmistusta, ja yhä useammat kielentutkijat julkaisevatkin aineistonsa avoimesti kaikkien tutkijoiden saataville esimerkiksi Kielipankin kautta.[9]

Tutkimusaineistosta tehtyjen havaintojen kuvaaminen ja tulkitseminen tarkoittaa samalla myös käytettyjen teoreettisten käsitteiden ja näkökulmien kytkemistä konkreettiseen materiaaliin. Tutkimuksen teoriatausta auttaa havaintojen systematisoinnissa ja rikastuttaa tutkijan ilmiötä koskevia pohdintoja tarjoamalla välineitä käsitellä ilmiötä monipuolisemmin kuin mitä ilman tällaisia taustatietoja olisi mahdollista (Eskola & Suoranta 1999, 80; vrt. Creswell 2014, 64). Toisaalta tutkimusaineistosta tehtyjä havaintoja voidaan tarkastella myös teorian testauksen näkökulmasta ja pohtia, miltä osin ne tukevat aikaisempia

tutkimustuloksia ja miltä osin teoriaa olisi ehkä syytä tarkentaa. Tulosten tulkinnassa pyritään siis vastaamaan kahteen kysymykseen, joiden keskinäinen painoarvo voi totta kai vaihdella eri tutkimuksissa: Mitä käytetty tutkimusaineisto paljastaa tutkittavan kielen ominaisuuksista valitussa teoriataustassa tarkasteltuna? Miten käytetty tutkimusaineisto voi kehittää tai haastaa käytetyn teoriataustan näkemyksiä? Ilman tällaista tutkimustulosten ja teoriataustan vuoropuhelua tutkimusaineistosta tehdyt päätelmät ovat vain irrallinen joukko havaintoja, joita ei kytketä mihinkään aiempaan tutkimukseen, eikä niiden arvoa ole näin ollen mahdollista arvioida. Tutkimustulosten suhteuttaminen laajempaan tieteelliseen kontekstiin onkin olennainen osa tutkimuksen raportointia.

Valittu teoreettinen näkökulma ohjaa tulosten tulkinnan suuntaa, ja eri näkökulmien käyttäminen tarjoaa mahdollisuuden käyttää samaa tutkimusaineistoa useampaan eri tarkasteluun. Esimerkiksi edellä esitelty grammitietokanta on rakennettu niin, että sen teoriataustana on kognitiivinen kielentutkimus. Tämän teoriakehyksen sisällä tietokanta on toiminut materiaalina paitsi grammien syntaktis-semanttisessa analyysissa myös niin sanottujen viittauskehysten tarkastelussa[10] ja kieliopillistumistutkimuksessa, ja koko ajan tärkeänä taustana on toiminut myös kielitypologinen ajattelu eli se, mitä vastaavista ilmiöistä maailman kielten näkökulmasta tiedetään. Valittu teoria voi luonnollisesti tuoda mukanaan myös tarpeen liittää tutkimukseen muitakin menetelmiä: esimerkiksi kieliopillistumistutkimus edellyttää kielihistoriallisten menetelmien hyödyntämistä (ks. Historiallinen ja vertaileva kielentutkimus tk.).

Nykyisin on yhä yleisempää, että aineistopohjaisessa kielentutkimuksessa yhdistetään sekä laadullista että määrällistä tutkimusta. Laadullisia menetelmiä hyödyntäen voidaan esimerkiksi ensin kartoittaa ilmiötä kokonaisvaltaisesti ja nostaa sitten kiinnostavia kohtia määrälliseen tarkasteluun ilmiön yleisyyden kartoittamiseksi. Toisaalta ilmiötä voidaan myös tarkastella ensin määrällisesti ja poimia määrällisessä tarkastelussa huomiota herättäviä kohtia tarkempaan laadulliseen tarkasteluun. Laadullinen aineistopohjainen kielentutkimus ja määrällinen korpustutkimus tulisikin mielestämme nähdä toisiaan täydentävinä menetelminä, jotka soveltuvat erilaisiin tarkoituksiin ja vastaavat erilaisiin tutkimuskysymyksiin. (Ks. Johdannon luvut 1 ja 4 tk.)

# 5. Yhteenveto

Laadullisen aineistopohjaisen kielentutkimuksen etuja ovat sen hyvä soveltuvuus erilaisiin tutkimusaiheisiin ja menetelmän joustavuus. Se soveltuukin lähes kaikenlaisten tutkimusaineistojen analysointiin. Sen erityisiä vahvuuksia ovat monimutkaisten ilmiöiden jäsentäminen, näkökulmien laajentaminen ja ymmärryksen lisääminen tutkimuskohteesta, joka voi olla entuudestaan kartoittamaton. Yleisesti laadullisen analyysin heikkouksina pidetään tutkimuksen hitautta, työläyttä ja huonoa yleistettävyyttä, etenkin jos käytetään pieniä tutkimusaineistoja. Kielentutkijoiden saatavilla on kuitenkin yhä enemmän tutkimusaineiston käsittelyä helpottavia ja nopeuttavia automaattisia työkaluja, joiden ansiosta yhä suurempien tutkimusaineistojen laadullinen analyysi tulee mahdolliseksi. Lisäksi tutkimustulosten luotettavuuden varmistamiseksi on kehitetty erilaisia menetelmiä, kuten tutkimusaineiston systemaattista järjestämistä tukevia ohjelmia ja kahden koodaajan käyttöä koodauksen laadunvarmistuksessa.

**Aiheesta lisää:**

Denzin, Norman K. & Lincoln, Yvonna S. (toim.). 2005. *Sage Handbook of Qualitative Research*. Thousand Oaks (CA): Sage.

Dörnyei, Zoltán. 2007. *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. Oxford: Oxford University Press.

Hasko, Victoria. 2012. Qualitative corpus analysis. Julkaisussa: Chapelle, Carol A. (toim.) *The Encyclopedia of Applied Linguistics*. Blackwell Publishing 2013– (jatkuva julkaisu).

## VIITTEET

1 Kiitämme artikkelimme toimittajia, kommentoijia ja arvioijia hyödyllisistä kommenteista. Milla Luodonpää-Manni on saanut rahoitusta artikkelin kirjoittamiseen Varsinais-Suomen rahastolta ja Krista Ojutkangas Suomen Akatemialta (285739).

2 Keskeinen osa fennistiikan syntyä on tietysti myös kielioppien kirjoittaminen, mikä oli tärkeää mm. siksi, että suuri osa vielä 1800-luvun kielimiehistä oli äidinkieleltään tai vähintään koulusivistykseltään ruotsinkielisiä (ks. esim. Häkkinen 2008, 65–73).

3 https://www.kotus.fi/kielitieto/kielenkayton_lajit/puhuttu_kieli/murteenseuruuhanke.

4 Hankkeen kuvaus: https://www.utu.fi/fi/yliopisto/humanistinen-tiedekunta/suomen-kieli-ja-suomalais-ugrilainen-kielentutkimus/suomen-kieli/tutkimus. Aineisto kielipankissa: http://urn.fi/urn:nbn:fi:lb-2014073026

5 http://syntaxarchives.utu.fi/Graduja/.

6 Eri menetelmiä voi olla tarpeen yhdistää myös silloin, jos kaikkia olennaisia tutkimuskysymyksiä ei pysty ratkaisemaan aitojen kielenkäyttöaineistojen avulla (esim. Luodonpää-Manni (2016) aineistopohjaisen laadullisen kielentutkimuksen menetelmän ja kyselymenetelmän yhdistämisestä).

7 Lauseopin arkiston esittely osoitteessa http://syntaxarchives.suo.utu.fi/. Arkiston aineisto on käytettävissä Korpissa (https://www.kielipankki.fi/tuki/korp/).

8 Tätä vaihetta tutkimusaineiston käsittelyssä nimitetään *annotoinniksi*, mikäli se tapahtuu niin, että merkinnät ovat konelukuisessa muodossa (esim. xml). Koodaamiseen sisältyy siis käsityön ajatus.

9 Ks. https://avointiede.fi/fi.

10 Viittauskehyksissä on kyse siitä, miten ns. akseligrammien ilmaisemat sijainnit tulkitaan. Esimerkiksi lause *Pallo on auton edessä* voidaan tulkita joko niin, että pallo on tilanteen tarkastelijan ja auton välissä (ns. relatiivisen viittauskehyksen mukainen tulkinta), tai niin, että pallo on auton edessä auton kulkusuuntaan nähden (ns. luontaisen viittauskehyksen mukainen tulkinta). (Levinson 2003; suomesta Ojutkangas 2005.)

## KIRJALLISUUS

Creswell, John W. 2014. *Research Design. Qualitative, Quantitative and Mixed Methods Approaches*. Los Angeles (CA): Sage.

Dixon, Robert. M. W. 2010. *Basic linguistic theory. Vol. 1. Methodology*. Oxford: Oxford University Press.

Dozat, Timothy, Qi, Peng & Manning, Christopher D. 2017. *Stanford's Graph-based Neural Dependency Parser at the CoNLL 2017. Shared Task*. [verkkoaineisto]. [viitattu 18.5.2018]. Saatavissa: https://www.aclweb.org/anthology/K17-3002/.

Eskola, Jari & Suoranta, Juha. 1999. *Johdatus laadulliseen tutkimukseen*. Tampere: Vastapaino.

Firth, John R. 1968. *Selected Papers of J. R. Firth 1952–59*. Toimittanut F. R. Palmer. London: Longmans.

Frege, Gottlob. 1884. *Die Grundlagen der Arithmetik: Eine logisch-mathematische Untersuchung* über den Begriff der Zahl. [verkkoaineisto]. [viitattu 19.3.2020]. Breslau: Verlag von Wilhem Koebner. Saatavissa: https://archive.org/details/diegrundlagendeoofreggoog.
Glynn, Dylan. 2014. Polysemy and synonymy: Cognitive theory and corpus method. Julkaisussa: Glynn, Dylan & Robinson, Justyna A. (toim.) *Corpus Methods for Semantics. Quantitative Studies in Polysemy and Synonymy*. Amsterdam: John Benjamins, 7–38.
Hasko, Victoria. 2012. Qualitative corpus analysis. Julkaisussa: Chapelle, Carol A. (toim.) *The Encyclopedia of Applied Linguistics*. Blackwell Publishing 2013– (jatkuva julkaisu).
Haverinen, Katri, Nyblom, Jenna, Viljanen, Timo, Laippala, Veronika, Kohonen, Samuel, Missilä, Anna, Ojala, Stina, Salakoski, Tapio & Ginter, Filip. 2014. Building the essential resources for Finnish: the Turku Dependency Treebank. *Language Resources and Evaluation* 48:3, 493–531.
Hovdhaugen, Even, Karlsson, Fred, Henriksen, Carol & Sigurd, Bengt. 2000. *The History of Linguistics in the Nordic Countries*. Jyväskylä: Societas Scientiarum Fennica.
Häkkinen, Kaisa. 1994. *Agricolasta nykykieleen. Suomen kirjakielen historia*. Helsinki: WSOY.
Häkkinen, Kaisa. 2008. *Suomen kielen historia 2: Suomen kielen tutkimuksen historia*. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 78. Turku: Turun yliopisto.
Ikola, Osmo (toim.). 1985. *Lauseopin arkiston opas*. Lauseopin arkiston julkaisuja 1. Turku: Turun yliopisto.
Inaba, Nobufumi. 2007. Mikael Agricolan teokset tietokannan muodossa. Julkaisussa: Häkkinen, Kaisa & Vaittinen, Tanja (toim.) *Agricolan aika*. Helsinki: BTJ Kustannus, 147–162.
Jantunen, Jarmo Harri. 2009. Ei pelkästään mielikuvituksen puutteen vuoksi – Kieliaineistojen systemaattinen käyttö kielentutkimuksessa. *Virittäjä* 113:1, 101–113.
Jääskeläinen, Anni. 2013. *Todisteena äänen kuva. Suomen kielen imitatiivikonstruktiot*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-8860-5.
Karlsson, Fred. 1975. Fennistiikan tieteenparadigmasta ja sen ohjausvaikutuksesta. *Virittäjä* 79:2, 179–193.
Karttunen, Mia & Rouhikoski, Anu. 2010. Kentällä kokeillen: murresyntaksin tutkijat uusia aineistoja kokoamassa. *Virittäjä* 114:3, 426–432.
Leino, Pentti, Majapuro, Marketta & Kristiansson-Seppälä, Anna-Liisa (toim.). 1984. *Fennistiikan opinnäytteitä 2. Bibliografinen luettelo Suomessa vuosina 1976–1983 tehdyistä suomen kieltä ja sen sukukieliä koskevista painamattomista akateemisista opinnäytteistä*. Helsinki: Suomalais-Ugrilainen Seura.
Levinson, Stephen C. 2003. *Space in Language and Cognition. Explorations in Linguistic Diversity*. Cambridge: Cambridge University Press.
Luodonpää-Manni, Milla. 2016. *'Théorie', 'hypothèse', 'modèle' et 'méthode' dans le domaine de la linguistique: Perspectives sémasiologique et onomasiologique*. Väitöskirja. Turku: Turun yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-29-6490-1.
Luodonpää-Manni, Milla, Penttilä, Esa & Viimaranta, Johanna. 2017. Introduction. Julkaisussa: Luodonpää-Manni, Milla, Penttilä, Esa & Viimaranta, Johanna (toim.)

*Empirical Approaches to Cognitive Linguistics: Analyzing Real-Life Data*. Newcastle Upon Tyne: Cambridge Scholars Publishing, 1–21.
Lyytikäinen, Erkki. 1984. Suomen kielen nauhoitearkiston neljännesvuosisata. *Virittäjä* 88:4, 449–457.
Lyytikäinen, Erkki & Yli-Paavola, Jaakko. 2010. Suomen kielen nauhoitearkisto 50-vuotias. *Virittäjä* 114:3, 411–419.
Niiniluoto, Ilkka. 1980. *Johdatus tieteen filosofiaan – käsitteen- ja teorianmuodostus*. Helsinki: Otava.
Nummila, Kirsi-Maria. 2012. *Tekijännimet Mikael Agricolan teosten kielessä: henkilötarkoitteisten johdosten merkitykset, funktiot ja rakenteet*. Annales Universitatis Turkuensis C 328. Väitöskirja. Turku: Turun yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-29-4866-6.
Ojutkangas, Krista. 2005. Viittauskehykset ja tarkastelunäkökulma – miten sijaintia perusakselilla kuvataan? *Virittäjä* 109:4, 525–551.
Ojutkangas, Krista. 2008a. Mihin suomessa tarvitaan *sisä*-grammeja? *Virittäjä* 2008:3, 382–400.
Ojutkangas, Krista. 2008b. Kielentutkijan mielikuvituksen puute eli mihin todelliseen kielenkäyttöön perustuvaa aineistoa tarvitaan. *Virittäjä* 112:2, 276–284.
Ojutkangas, Krista. 2010. Eläimellistä menoa: tutkimus kahdesta eläimen nimestä koostuvista lajinnimistä. Julkaisussa: *Sanoista kirjakieliin. Juhlakirja Kaisa Häkkiselle 17. marraskuuta 2010*. Suomalais-Ugrilaisen Seuran Toimituksia = Mémoires de la Société Finno-Ougrienne 259. Helsinki: Suomalais-Ugrilainen Seura, 203–218.
Pajunen, Anneli. 2001. *Argumenttirakenne: asiaintilojen luokitus ja verbien käyttäytyminen suomen kielessä*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Payne, Thomas. 1997. *Describing Morphosyntax. A Guide for Field Linguist*. Cambridge: Cambridge University Press.
Päiviö, Pia-Maria. 2007. *Suomen kielen* asti *ja* saakka. *Terminatiivisten partikkelien synonymia, merkitys, käyttö ja kehitys sekä asema kieliopissa*. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 75. Turku: Turun yliopisto.
Ravila, Paavo. 1967. *Totuus ja metodi*. Helsinki: WSOY.
Suutari, Toni. 2010. Suomen kielen nauhoitearkisto – vireä viisikymppinen. *Virittäjä* 114:3, 423–426.
Svorou, Soteria. 1993. *The Grammar of Space*. Amsterdam: John Benjamins.
Tissari, Heli. 2017. Corpus-linguistic approaches to metaphor analysis. Julkaisussa: Semino, Elena & Demjén, Zsófia (toim.) *The Routledge Handbook of Metaphor and Language*. Milton Park: Routledge, 117–130.
Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
Tummers, José, Heylen, Kris & Geeraerts, Dirk. 2005. Usage-based approaches in Cognitive Linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory* 1:2, 225–261. DOI: 10.1515/cllt.2005.1.2.225.
Vartiainen, Turo. 2016. *Challenges in Categorization: Corpus-based Studies of Adjectival Premodifiers in English*. Mémoires de la Société Néophilologique de Helsinki. Tome CI. Helsinki: Société Néophilologique. [verkkoaineisto]. [viitattu 19.3.2020]. Saatavissa: http://urn.fi/URN:ISBN:978-951-9040-58-5.

# Luku 8 Käsiteanalyysi kielentutkimuksessa

*Reetta Konstenius*

**Mikä?**

Käsiteanalyysi tarkoittaa usein jonkin tietyn (teoreettisen) käsitteen merkityksen purkamista ja pilkkomista sekä suhteuttamista lähikäsitteisiin ja johonkin laajempaan taustakäsitejärjestelmään. Käsiteanalyysin päämääränä on koherentti ja yhtenäinen kuva jonkin tietyn käsitteen merkityssisällöstä ja suhteista muihin käsitteisiin. Se on tieteellinen filosofinen menetelmä, jolla voidaan jäsentää abstraktien käsitteiden tietosisältöjä. Se on myös välttämätön osa kaikkea tieteellistä tutkimusta, sillä tutkijoiden on tutkimuksessaan määriteltävä, avattava ja perusteltava käyttämänsä keskeiset käsitteet.

**Katso myös:**

Luku 1 Kielitaju kielentutkijan työkaluna

Luku 3 Haastattelu aineistonkeruun metodina sosiolingvistiikassa

# 1. Johdanto

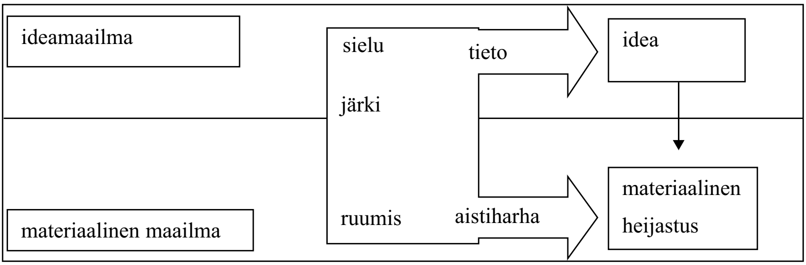
Käsiteanalyysi tarkoittaa nimensä mukaisesti käsitteiden eli ajattelun ja päättelyn perustana toimivien merkityssisältöjen analyysia, ja se eroaa siksi muista vastaavanlaisista metodeista, kuten sanasemantiikasta, leksikologiasta tai terminologiaopista. Käsiteanalyysin päämääränä on koherentti ja yhtenäinen kuva jonkin tietyn käsitteen merkityssisällöstä. Sanasemantiikka vuorostaan on leksikologian osa, jossa tutkitaan sitä, mitä eri merkityksiä, suhteita tai muutoksia lekseemeillä on. Vaikka nämä kaksi tutkimusmetodia käyttävät jokseenkin samankaltaista menettelyä, metodien väliset erot ilmenevät erityisesti tutkimusnäkökulmassa ja tutkimuksen päämäärissä. Lekseemit kuuluvat kielen sanastoon, ja ne toteutuvat erilaisina sananmuotoina. Sanasemantiikassa tutkimuksen kohde on yleensä jollakin tavoin merkityksen ja muodon suhde. Sanasemantiikassa tärkeitä ovat sellaiset termit, jotka ilmaisevat leksikaalisia suhteita, kuten *polysemia*, *metonymia*, *homonymia* ja *antonymia*. Käsitteet vuorostaan toteutuvat merkityksissä, ja käsiteanalyysin päämääränä on eritellä ja konstruoida merkityksiä suhteessa johonkin taustakäsitejärjestelmään. Käsiteanalyysi eroaa siten myös onomasiologisesta kielentutkimuksesta, jossa tutkitaan jonkin tietyn käsitteen ilmenemistä eri murteiden tai kielten sanastossa (esim. Nirvi 1952). Käsiteteanalyysin erityispiirteenä voidaan pitää sitä, että tutkimus suuntautuu yleensä yhteen tiettyyn käsitteeseen (ks. esim. Jouhki & Pennanen 2016; Hyvärinen, Kurunmäki, Palonen, Pulkkinen & Stenius 2003).

# 2. Mitä tutkitaan, kun tutkitaan käsitettä?

Kielentutkijan vastaus kysymykseen, mitä käsitteet ovat, määrittää sekä analyysitekniikkaa että tutkimuksen taustafilosofiaa. Erityisen mielenkiintoista käsitteiden tutkimuksen näkökulmasta on se, että teoria käsitteistä johtaa yleensä sellaisiin metafyysisiin sitoumuksiin, joilla on

kokonaisvaltaisia seurauksia tutkimuksen edustamalle ja tuottamalle maailmankatsomukselle. Seuraavassa on muutamia tunnettuja esimerkkejä tästä.

Käsiterealisti Platon ajatteli, että ideat ovat ikuisia universaaleja ja siksi käsitteiden tutkimus kohdistuu johonkin reaalitodellisuuttakin todellisempaan transendentaaliseen ilmiöön. Käsiteanalyysin avulla saavutettu tutkimustulos kertoo itse todellisuuden perimmäisistä rakenteista. Platonin käsiterealismi on dualismia, sillä se olettaa kaksi todellisuutta: idea- ja materiamaailman. Materiamaailmassa elävä ihminen saavuttaa ikuiset ideat sielunsa avulla. Tämä sielussa tapahtuva muistaminen on rationalistin mukaan todellista tietoa, kun taas aistein havaittu tieto on harhaa (ks. kuvio 1).

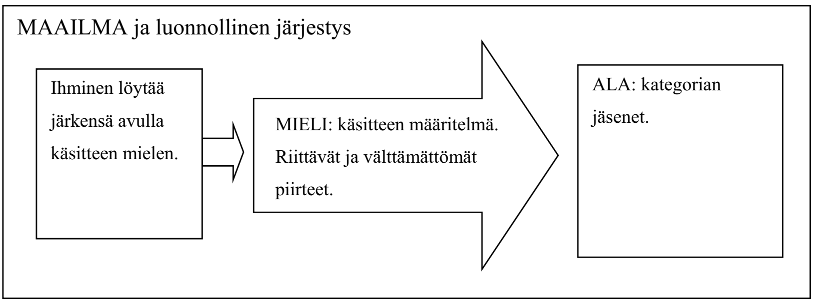



Kuvio 1. Platonin dualistinen käsiterealismi.

Ideaoppi voi olla myös monistinen: esimerkiksi jotkin uudet modernit metafyysiset teoriat ovat ehdottaneet, että maailman perimmäinen luonne on matemaattinen. Universumit (engl. *multiverse*) perustuvat matemaattiselle rakenteelle (engl. *mathematical universe hypothesis*) ja ovat luonteeltaan komputationaalisia (engl. *computable universe hypothesis*). Havaitsemamme ulkoinen todellisuus on siten fysikaalista komputaatiota ja ihminen on itsestään tietoinen komputationaalinen alarakenne (engl. *self-aware substructure*) (ks. esim. Tegmark 2014). Yleiskäsitteet ovat tämän näkemyksen mukaan matemaattisia.

Maltilliset käsiterealistit seuraavat Aristoteleen (1994) jäljissä ja yleensä uskovat, että käsitteet tai universaalit ovat osa maailman loogista rakennetta. Käsitteiden analyysin tulos siis paljastaa jotakin maailman

rakenteesta. Tämän klassisen näkemyksen mukaan käsitteet ovat tarkkarajaisia ja ne voidaan määritellä **välttämättömien** ja **riittävien** piirteiden kautta. Klassinen käsiteanalyysi onkin loogista ja strukturoitua välttämättömien ja riittävien piirteiden määrittelemistä. Klassisessa teoriassa käsitteen määritelmää kutsutaan usein **mieleksi** tai **intensioksi** ja vuorostaan **alaksi** tai **ekstensioksi** sitä joukkoa, johon kuuluvat määritellyn kategorian jäsenet. Mitä laajempi on käsitteen ala, sitä vähemmän yhteistä kategoriaan kuuluvilla olioilla yleensä on, eli käsitteen mieli eli intensio supistuu. Esimerkiksi käsite ELÄIN kattaa monia keskenään erilaisia olioita, ja sen määritelmä on siten lavea. Vastaavasti KOIRA kattaa kapeamman alan olioita, ja tämän käsitteen määritelmä on näin yksityiskohtaisempi (Kakkuri-Knuuttila 1998).[1] Käsiterealismille on tyypillistä jonkinlainen uskomus siihen, että käsitteet kuvastavat maailman luonnollisia kategorioita. Syy siihen, miksi kategorioita IHMINEN ja KOIRA yhdistää kategoria ELÄIN, on se, että sama luonnollinen selitys koskee kumpaakin kategoriaa. Käsitteiden ja maailman välillä on jonkinlainen välttämätön viittaussuhde: ihminen havaitsee tai löytää kategorian jäsenten väliset suhteet järkensä avulla.




Kuvio 2. Aristoteelinen käsitys käsitteiden luonteesta.

Aristoteeliseen perinteeseen voidaan lukea erilaiset totuusehtosemantiikat ja referenssiin perustuvat merkitysteoriat, sillä ne olettavat merkityksen rakentuvan kielen ja maailman välisessä suhteessa (Forrai 2001). Sekä Platon että Aristoteles olivat käsiterealisteja, eli he uskoivat, että käsitteillä on itsenäinen ihmisestä riippumaton olemassaolo.

Epistemologia kuitenkin erottaa nämä kaksi näkemystä: Platon oli rationalisti, sillä hän uskoi, että sielu havaitsee käsitteen totuuden. Aristotelesta vuorostaan voidaan pitää historian ensimmäisenä tunnettuna empiristinä, sillä hän uskoi, että järki näkee reaalimaailmassa esiintyvissä yksilöjoukossa yhteiset tekijät.

Käsiterealistit voidaan siis jakaa rationalisteihin ja empiristeihin. Esimerkiksi formaalin kielentutkimuksen semantiikan taustateorioista Jerry Fodorin (1981) *mentalese*, jonka mukaan ajattelun kielellä on kielioppia muistuttava rakenne, edustaa rationaalista käsiterealismia. Fodor pyrki laajentamaan Jerrold Katzin kanssa Chomskyn ajatuksia semantiikasta, ja yhdessä he julkaisivat artikkelin *The Structure of a Semantic Theory* (1963). Fodorin teoriaa voi olla monen vaikea sulattaa, sillä teesi käsitteiden sisäsyntyisyydestä johtaa siihen väitteeseen, että ihmiset kautta aikojen olisivat syntyneet esimerkiksi sellaiset käsitteet kuin INTERNET ja KOGNITIIVINEN KIELIOPPI mielessään.

Siinä missä klassisessa teoriassa lähdetään liikkeelle määritelmästä, josta dedusoidaan kategorian jäsenet, uudemmissa merkitysteorioissa edetään vastakkaisessa suunnassa. Esimerkiksi kielen käyttöteorioissa ja prototyyppiteorioissa päätelmät käsitteen merkityssisällöstä johdetaan kategorian jäsenien yhteisistä piirteistä induktiivisesti, alhaalta ylös, eli jossakin kerätyssä aineistossa esiintyvistä viittauskohteista päätellään vaikkapa käsitteen ELÄIN keskeiset tai yhdistävät piirteet (ks. esim. Taylor 2008; Geeraerts 2006). Modernin käsiteanalyysin taustalla on usein jonkinlainen versio Wittgensteinin (1953) kielen käyttöteoriasta: merkitys muodostuu käyttöyhteyden kautta. Wittgenstein käyttää termiä *kielipeli*, jossa pelin säännöt ja itse pelattu peli muodostavat merkityssysteemin. Kielen käyttöteoriassa jonkin lekseemin ja sen synonyymien esiintymiä kerätään ja tulkitaan käyttöyhteydessään rinnakkain. Kielen käyttöteoriat ymmärtävät kielen merkityksen luonteeltaan sosiaaliseksi. Tämä tarkoittaa sitä, että vaikka jollakin olisi väärä idea sanan *eläin* merkityksestä, tämä harhaluulo ei millään tavalla vaikuta kielen merkityksen muodostumiseen eli muuta sanan vakiintunutta merkitystä.

Kielen käyttöteorioiden ja prototyyppiteorioiden välinen ero on juuri se merkitys, joka annetaan ihmisen mielelle. Kognitiivinen kielitiede, jossa prototyyppiteoriaa on sovellettu, esimerkiksi olettaa, että käsitteet

ja kategoriat ovat ihmisen mielessä, ja niitä kutsutaan vaihtelevasti *neurologisiksi kuvioiksi*, *mielikuviksi* tai *prototyypeiksi* (ks. esim. Dodge & Lakoff 2005; Kenttä 2001). Tämä koulukunta antaa merkitystä muodostavan roolin ihmisen psykologiselle kyvylle kategorisoida, ja tutkimus kohdistuu usein juuri kategorian rakenteeseen ja sen jäseniin. Lakoff (1987) esittää, että ihmisen havaintomekanismit yhdessä kulttuuristen arvojen ja käytäntöjen kanssa ohjaavat kategoriapäätelmiin siten, että kategorian yläpäässä voi toisinaan olla melko mielivaltainenkin abstraktio. Esimerkiksi käsitteitä NAINEN ja TULI voi yhdistää jossakin kielessä yläkategoria VAARA.

Edellisen perusteella huomaamme, että se, nimitetäänkö käsitteen merkityssisältöä *mieleksi*, *mielikuvaksi* vai *käytöksi*, riippuu omaksutuista merkitysnäkemyksistä. Yhteistä useimmille merkitysteorioille näistä eroista huolimatta on, että abstraktien ja vertailevien käsitteiden oletetaan olevan luonteeltaan erilaisia tai pitävän sisällään yleensä muita käsitteitä. Abstraktit käsitteet eivät välttämättä viittaa mihinkään omaan luokkaan reaalimaailmassa; emme voi havaita aisteillamme tai ajatella esimerkiksi negaatiota (EI) tai vertailusuhdetta (ISOMPI) erillään muista käsitteistä (ks. esim. Lukose, Delugach, Keeler, Searle & Sowa 1997). Konkreettisten käsitteiden (esim. HEVONEN) ajatellaan vuorostaan nimeävän tai edustavan todellisuudessa havaitsemiemme luonnollisten luokkien tai kategorioiden yleisiä piirteitä.

Merkitysteoriat voidaan edelleen jakaa kontekstuaalisiin (kielen käyttöteoriat) ja atomistisiin (prototyyppiteoriat, totuusehtosemantiikka). Kun kontekstuaaliset teoriat korostavat asiayhteyksiä, niin atomistiset teoriat olettavat, että käsitteellä on käyttöön nähden tietyssä määrin itsenäinen merkityssisältö.

Merkitysteorioiden suuressa tarjonnassa uudet teknologiat tarjoavat kielentutkijalle pragmaattisen tavan päätellä jotakin merkitysteorian pätevyydestä. Esimerkiksi uudet oppivat tekoälyt käyttävät hyväkseen hierarkkista merkitysteoriaa. *Convolutional neural networks* pilkkoo havainto- ja oppimispäätelmät hierarkkisiin rakenteisiin eli käsittelee kielellistä informaatiota käsitteenmuodostuksessa paloina, joita se tulkinnan kuluessa vertaa toisiinsa koherenssin muodostamiseksi. Tämä *deep learning* -algoritmi, jota käytetään yhä enemmän puheen tulkinnassa,

käännöksissä ja tekstin ymmärryksessä, on oppiva neuroverkko, joka muodostaa käsitteen alhaalta ylös ja vuorottelee havainnosta tehdyn induktion ja päättelyyn perustuvan deduktion välillä (LeCun & Bengio 1995). Tämä voi kertoa siitä, että myös ihmiset vuorottelevat erilaisten syötteiden ja prosessien välillä.

Kielifilosofiassa on siis esitetty keskenään suuri joukko erilaisia merkitysteorioita, joista osa esittää, että kielen merkitys on itsenäinen eli mikään ihmisen tulkinta ei vaikuta siihen, ja toiset taas esittävät, että viestin lähettäjä tai tulkitsija määrää merkityksen. Alla olevassa laatikossa on ajatuskoe, joka auttaa arvioimaan omia intuitioita kielen merkityksen rakentumisesta (ks. esim. Riemer 2015).

**Ajatuskoe käsitteestä *kieli***

Tämän ajatuskokeen tehtävänä on auttaa sinua arvioimaan, millä tavalla itse uskot, että viestijän intentio, tulkinta, tietoisuuden aste tai alkuperä vaikuttavat merkitykseen. Ajatuskokeen avulla voit arvioida intuitioitasi siitä, missä olosuhteissa sanalla on mielestäsi jokin tietty merkitys ja milloin ylipäätään kyse on kielestä, milloin muusta kuin kielestä.

Löydät lapun, jossa lukee *aita*. Suomeksi *aita* on ’maa-aluetta rajoittava rakenne’ mutta baskiksi ’isä’. Erilaisen tausta- ja tilannetiedon sekä tekemiesi havaintojen perusteella saatat päätyä mm. seuraaviin eri tulkintoihin:

a) Kirjoittaja on baski, joka kirjoitti lappuun *isä* baskiksi osoittaen sen isälleen.

b) Kirjoittaja on suomalainen, joka kirjoitti lappuun *aita* viitaten omaan puutarha-aitaansa.

c) Kirjoittaja on baski, joka kirjoitti lappuun *isä* baskiksi, mutta uskot, että sana on suomea ja tarkoittaa puutarha-aitaa.

d) Sanan *aita* kirjoitti baski narraavan suomalaisen sanelemana. Baski luuli kirjoittaneensa *isä*, vaikka sanelija mielessään tarkoitti puutarha-aitaa.

e) Teksti *aita* on osa revittyä lappua, jossa alun perin luki *asianlaita*.

f) Kirjainjonon tuotti kone, joka sattumalta yhdisti kirjaimet keskenään.

g) Kirjainjonon tuotti kone, joka puutarhamessuilla aivan tarkoituksella osoitti jokaiselle esineelle nimen.

h) Kirjainjonon tuotti kone, joka puutarhamessuilla aivan tarkoituksella osoitti jokaiselle esineelle nimen mutta sekoitti pelargoniat ja aidat keskenään. Kone nimesi vahingossa pelargoniat aidoiksi.

i) Suomalainen yritti kirjoittaa *auta*, mutta jostain syystä *u* näyttää *i*:ltä.

j) Lappu on osa arkeologista löytöä, eikä ole olemassa enää yhtään kieliyhteisön jäsentä, joka osaisi kertoa, mitä *aita* kyseisellä ja tilanteessa tuntemattomalla kielellä tarkoittaa.

Onko kieli siis riippuvainen puhujan intentiosta, kuulijan tai lukijan tulkinnasta, ihmisen ideoivasta mielestä? Tuottaako mitään ymmärtämätön kone kieltä? Voiko virheellinen tulkinta viestistä olla edelleen kieltä? Voiko virheellinen tietokoneohjelmisto tuottaa vahingossa kieltä? Onko kaikki kieli kommunikaatiota? Oleellista tässä ajatuskokeessa on ymmärtää, että pelkkä kirjoitusmuoto tai kirjainjono ei vielä oikeastaan kerro, kuuluuko se johonkin kieleen tai onko se kieltä. Vaikka muoto olisi eräs edustus lekseemistä, se ei silloinkaan välttämättä ehkä olisi kaikissa tapauksissa kieltä, jos kieleen kuuluu intentio, oikea tulkinta tai kieliyhteisön normi.

## 3. Millaisessa tutkimuksessa käsiteanalyysia käytetään?

Käsiteanalyysia pidetään usein filosofialle tyypillisenä menetelmänä, sillä se on luonteeltaan analyyttista ja sen päämääränä on synteettinen yhtenäinen tiedollinen esitys käsitteen sisällöstä eli **näkemys**. Esimerkiksi kielifilosofiassa tutkitaan kieleen liittyviä käsitteitä ja niiden sisäistä rakennetta sekä yhteyttä muihin käsitteisiin ja teorioihin. Kielifilosofia pohtii sellaisia kysymyksiä kuin ”Mitä on merkitys?” ja ”Mikä on normi?”. Usein käsiteanalyysi on ainoa tieteellinen tutkimusmetodi, jolla voidaan jäsentää abstraktien käsitteiden tietosisältöjä, koska ne pitävät sisällään ainoastaan muita käsitteitä eivätkä viittaa mihinkään aistein havaittavaan luonnolliseen luokkaan reaalimaailmassa (Rescher 2001).

Käsiteanalyysi ei rajoitu kuitenkaan pelkästään filosofiaan, vaan se on välttämätön osa kaikkea tieteellistä tutkimusta; tutkijoiden on tutkimuksessaan määriteltävä käyttämänsä keskeiset käsitteet. Tieteen kielen tulee olla yksimerkityksistä ja täsmällistä, ja siksi käsitteiden huolellinen määrittely edellyttää käsitteen aikaisemman käytön sekä käsitteelle annettavien erilaisten merkitysten analysoimista. Tutkijan on eksplisiittisesti perusteltava käyttämänsä merkitykset ja osoitettava käyttämiensä käsitteiden ja muiden samankaltaisten käsitteiden väliset yhtäläisyydet ja erot (ks. Niiniluoto 1997; Kakkuri-Knuuttila 1998).

Erityisen oleellinen menetelmä käsiteanalyysi on kielitieteen metateoriassa, jossa tutkimuksen kohteena ovat erilaiset kieliteoriat, keskenään kilpailevat ja koulukuntakiistoja synnyttävät näkemykset tai jokin akateeminen keskustelu esimerkiksi jostakin valitusta keskeiskäsitteestä (ks. esim. Hamunen, Nieminen, Kelomäki & Dufva 2019). Kun selvitetään tietyn kieliryhmän, aikakauden, tekstilajin tai ajattelijan käyttämiä merkityksiä, käsitteiden analyysi, vertailu ja määrittely on usein päämetodi. Käsitteiden alkuperän, muutosten, historian, kielellisen alkuperän ja tekstilajittaisen variaation tarkastelu kuuluu kielitieteeseen, ja käytössä on silloin usein käsiteanalyysi. Käsiteanalyysia käytetään myös semantiikassa silloin, kun tutkimuksen kohteena on yksi koherentti tekstikokonaisuus, jonka sisältöä analysoimalla voidaan konstruoida

käytetyn käsitteen tai käsitteistön merkityssisältö. Myös diskurssianalyysissa ja leksikaalisessa semantiikassa voidaan toisinaan soveltaa käsiteanalyysia, kun selvitellään jonkin tietyn käsitteen tarkoitusta käyttöyhteyden valossa.

## 4. Käsiteanalyysin tavoitteet tutkimuksessa

Tutkimuksessa käsiteanalyysi voi olla useassa roolissa. Esimerkiksi haastattelututkimuksessa voidaan käsiteanalyysia käyttämällä abstrahoida vastauksista uusi käsite (ks. Haastattelu aineistonkeruun metodina sosiolingvistiikassa tässä kirjassa). Käsite voi siten olla **tutkimuksen tulos**. Tällöin ajatellaan, että käsiteanalyysin tavoitteena on reaalimääritelmä, sillä sen tavoite on kuvata reaalitodellisuutta ja tutkimuksen tulosta tarjotaan muille tutkijoille kehitettäväksi ja testattavaksi. **Reaalimääritelmä** koskee faktuaalisia seikkoja, ja se eroaa **nominaalimääritelmästä**, joka käsittelee verbaalisia sopimuksia. Reaalimääritelmä voi olla tosi tai epätosi, nominaalimääritelmä ainoastaan enemmän tai vähemmän käyttökelpoinen.

Tietty käsitteen määritelmä voi olla myös **hypoteesi**, jota tutkimuksen kuluessa testataan, kehitetään tai jäsennetään aineiston valossa. Tällöin tutkimuksen tavoitteena on usein käsitteen **deskriptiivinen määritelmä** eli yksityiskohtainen kuvaus käsitteen moninaisesta käytöstä. Päämääränä on kartoittaa käsitteen yhteyksiä muihin käsitteisiin, teksteihin ja teorioihin.

Deskriptiiviselle käsitetutkimukselle vastakkainen tavoite on yleensä kokeellisella tutkimuksella, jossa käsiteanalyysia käytetään rikkaan määritelmän supistamiseen: tavoitteena on tuottaa havaintotermein lausuttu kausaalinen ja semanttisesti köyhä **operationaalinen määritelmä**, jota voidaan käyttää kokeena tai josta voidaan konstruoida kokeellinen tutkimusasetelma. Havaintotermi viittaa välittömästi havaittaviin ominaisuuksiin tai suhteisiin, ja niiden avulla yleensä määritellään tutkimuksessa mitattavat yksiköt (Konstenius 2014). Jako havaittaviin ja

abstrakteihin termeihin voidaan ymmärtää myös Searlen (1995) erottelun avulla: värikäs paperilappu on havainnon ulottuvilla eli karkea fakta (engl. *brute fact*), kun taas tulkinta, että paperilappunen on rahaa ja sillä on arvo, on institutionaalinen fakta (engl. *institutional fact*).

Filosofisen käsiteanalyysin päämääränä on vuorostaan koetella käsitettä tai **falsifioida** analyysin kohteena oleva käsite. Tässä kehyksessä pyritään hylkäämään tai vahvistamaan käsitteitä niiden loogisten ominaisuuksien perusteella. Käsitettä koskevien loogisuusvaatimusten eli käsitteen taustaoletusten ja kategoriaoletusten on oltava keskenään ristiriidattomat. Sisäisesti ristiriitaiset käsitteet hylätään, koska niillä ei ole totuusarvoa. Totuusteoreettiset tavoitteet johtavat filosofisessa tutkimuksessa käyttämään käsiteanalyysin rinnalla informaalia ja formaalia logiikkaa, ajatuskokeita ja argumentaatioanalyysia.

# 5. Tutkimuksen vaiheet

Tässä luvussa selostetaan tiiviisti käsiteanalyyttisen menetelmän vaiheet ja kulku. Esitystä havainnollistetaan kuvitteellisella esimerkkitutkimuksella.

1. Tutkimuskohde
Käsiteanalyysi alkaa tutkimuskohteen eli tutkittavan käsitteen valinnalla.

> Esimerkkitutkimuksessa päätämme tutkia Eleanor Roschin (1973) käsitettä PROTOTYYPPI.

2. Näkökulma
Tutkimuskohteen valinnan jälkeen pohditaan, mikä voisi olla tutkimuksen anti tai se ajankohtainen tieteellinen kysymys, johon voidaan vastata tutkimalla valittua käsitettä.

> Lähestymme Roschin PROTOTYYPPI-käsitettä siitä näkökulmasta, onko kognitiivisen kieliopin määritelmä (erityisesti Langackerin [1987] tulkinta prototyypistä) uskollinen Roschin PROTOTYYPPI-käsitteelle.

3. Analyysin tavoitteiden määrittely
Kun tutkimuksen näkökulma on tarkennettu, on aika määritellä tutkimuksen tarkemmat tavoitteet. Tavoitteet voidaan muotoilla tutkimuskysymyksen muotoon.

> Esimerkkitutkimuksemme tavoitteena on tuottaa tuore näkemys siitä, millä tavoin kognitiivisen kieliopin terminologia on yhteneväinen Roschin teorian kanssa.

4. Aineiston valinta
Käsiteanalyysissa tarkastellaan useimmiten yhden kirjoittajan tekstiä tai jonkin jollain mittapuulla yhtenäisen ryhmän kirjoituksia kerrallaan, jotta rakentuisi koherentti käsitys käsitteestä. Käsite on aina osa laajempaa käsitejärjestelmää, minkä vuoksi käsitteen merkitys on ymmärrettävissä vasta, kun tutkimuksen kuluessa on selvitetty tutkitun käsitteen suhde taustakäsitejärjestelmän muihin käsitteisiin. Oleellista on valita sellainen yhtenäinen aineisto, jossa voidaan olettaa olevan yksi yhtenäinen käsitejärjestelmä tai joka edustaa tietyn kieliryhmän kielenkäyttöä. Käsiteanalyysissa syvyys ja tarkkuus ovat hyveitä – ei määrä!

> Koska esimerkkitutkimuksessamme tarkastelemme kognitiivisen kieliopin terminologiaa verrattuna Roschin teoriaan (erityisesti Langackerin [1987] tulkintaa prototyypistä suhteessa Roschin prototyyppi-käsitteeseen), tarvitsemme aineistoa molemmista teorioista. Valitsemme kohdeaineistoksi Langackerin teoksen *Foundations of Cognitive Grammar* (1987) sekä Roschin (1973–1999) relevantin tuotannon, jotta saamme kuvan PROTOTYYPPI-käsitteen muutoksista hänen kirjoituksissaan.

5. Tekstin esiymmärrys
Lähdetekstien hyvä tuntemus on käsiteanalyysin edellytys. Käsiteanalyysi pohjaa tietoon tekstin aihepiiristä ja tyylilajista sekä tieteellisestä koulukunnasta ja taustalla olevista tieteellisistä kiistoista, mikäli on kyse tieteellisestä tekstistä. Tekstin tausta ja ajassa käytetyt muut merkitykset muodostavat merkitysmatriisin valitun käsitteen rekonstruktiolle. Menetelmä perustuu siis viime kädessä tekstin tulkintaan, ja sillä on pyrittävä vastaamaan mm. seuraaviin kysymyksiin: Mihin kysymykseen teksti vastasi kirjoittajan aikalaisille? Mitä tutkittava käsite merkitsi aikalaisille arkikielessä tai tieteessä?

> Esimerkkitutkimuksessamme tämä tarkoittaa sitä, että kartoitetaan kognitiivisen kieliopin syntyhistoriaa ja ne formalismia vastustavat käsitykset, jotka koulukunta on esittänyt, pitäen mielessä formalismi–funktionalismi-eronteon dikotomisen rakenteen. Lisäksi kartoitetaan Roschin akateeminen konteksti psykologiassa (behaviorismi, kehityspsykologiset trendit jne.) sekä Wittgensteinin esittämän PERHEYHTÄLÄISYYS-käsitteen vaikutus Roschin tutkimusasetelmien arkkitehtuuriin.

6. Edeltävän epätietoisuuden tunnistaminen
Käsitteen merkityksen analyysi ja rekonstruktio edellyttävät käytettyjen ilmausten merkitysten tunnistamista ja ymmärtämistä aineistossa. Siksi on oleellista kiinnittää huomiota itselleen tuntemattomien merkitysten esiintymiseen tekstissä ja tutkimuksen kuluessa pohtia, kasvaako ymmärrys tekstin hankalista kohdista ja miten tuntemattomiksi jäävät merkitykset vaikuttavat luotettavaan käsitteen rekonstruktioon.

> Esimerkkitutkimuksessamme varmistamme tässä kohden, että ymmärrämme, mitä Rosch tarkoittaa peruskategorioilla.

7. Prosessin suunta
Käsitettä lähestytään tekstin sisäisestä käsitejärjestelmästä käsin, **hermeneuttisessa kehässä**, tulkiten merkityksiä aina laajempia taustakäsitejärjestelmiä vasten. Tekstistä haetaan tulkinta-apua käsitteen merkitys-

ten analyysiin ja rekonstruktioon. Tutkimuksen kuluessa ymmärrys tekstistä paranee.

8. Käsitteen esiintymien indeksointi teksti-, tilanne-, asia- ja käyttöyhteyksissään

**Esiintymä** (engl. *token*) on käsitteen konkreettinen esiintyminen jossakin kielellisessä yhteydessä. Käsitteen tehtävää ja roolia analysoidaan näissä teksti-, tilanne-, asia- ja käyttöyhteyksissä. Käytännössä käsitettä tulkitaan erikseen jokaisessa käyttöyhteydessään ja suhteessa asiayhteyteen, minkä jälkeen tehdään päätelmiä tekstiyhteydestä. Tarpeellista on myös koota ne yhteydet, joissa käsite ei esiinny mutta jotka ovat samankaltaisia kuin ne, joissa käsite esiintyy. Nämä voivat lisätä tietoa käsitteen merkityksestä tai kertoa kategorioiden rajoista.

> Havainnollistamme käsitevertailun rakennetta ja luonnetta poimimalla tarkasteluun esimerkkitutkimuksestamme yhden konkreettisen esiintymän, Langackerin (1987, 142) lauseen ”Certain types of nominals are clearly more prototypical than others”. Vertaamme lauseen tulkintaa Roschin vastaavan käsitteen määritelmiin. Jatkamme analyysia Langackerin *prototyyppi*-sanan ja sen johdosten esiintymälöydöksistä suhteessa Roschin tutkimuksiin.

9. Analyyttisten kysymysten esittäminen aineistosta (ks. myös Haarala 1981)

- Annetaanko käsitteelle reaali- tai nominaalimääritelmiä tekstissä?
- Liittyykö määrittelyyn havainnollisia esimerkkejä?
- Mitkä ovat käsitteen lähisynonyymit, ja mitä merkityseroja näillä on keskenään?
- Onko käsitteellä sukua tai yläluokkaa?
- Onko käsitteellä alaluokkia?
- Onko käsitteellä ominaisuuksia?
- Mikä tehtävä käsitteellä on asiayhteydessään?
- Onko käsitteellä tehtäviä ihmisten välisessä kommunikaatiossa? Kuuluuko käsite tietylle ryhmälle?
- Onko käsitteellä ajallisia ominaisuuksia tai prosessin osia? Onko

käsite itse osa jotakin prosessia tai kokonaisuutta?
- Millaisen todellisuuden käsite olettaa? Millaisia toimijoita, olemassaolon tasoja tai systeemejä se implikoi?
- Mitkä ovat käsitteeseen liittyvät muut käsitteet?
- Millä muilla käsitteillä tämän käsitteen ominaisuuksia, osia tai niiden puutteita kuvaillaan?

10. Synteesi

Analyysin tuloksista muodostetaan käsitteen kuvauksen synteesi, joka on argumenttien koherentti ja looginen kokonaisuus. Sen päämääränä on käsitettä koskeva teorianmuodostus. Kun käsitteen esiintymät on analysoitu, tulkintaa pyritään yhtenäistämään. Samaa merkitsevät ilmaukset pyritään sanomaan yhdellä ilmauksella. Käsitteiden tehtävänä on luokittelemalla ryhmittää olioiden yhteisiä piirteitä. Siksi käsiteanalyysin tulos eli käsitteen sisältö (intensio) on yleisten piirteiden joukko. On myös tarkasteltava, onko tekstissä käytetty käsite yksimerkityksinen, vertailemalla keskenään eri esiintymien analyysien tuloksia. Ovatko analysoitujen esiintymien kategoriaoletukset ja ontologinen taustateoria yhteneväiset eli loogisesti johdonmukaiset? Lausejoukko on ristiriitainen, jos kaikki lauseet eivät voi olla samanaikaisesti tosia. Mikäli osoittautuu, että käsitteellä on keskenään eriäviä taustaoletuksia, tulee pohtia, onko käsite sisäisesti ristiriitainen vai nimeääkö jokin termi paikoitellen eri käsitteitä. Tulee selvittää

- onko analyysin yhtenäisyyttä erottaville merkitystekijöille selityksiä tekstissä
- mitkä ovat esiintymien analyysien yhteiset ja mitkä erottavat merkitystekijät
- voiko yhdistävistä tekijöistä muodostaa käsite- tai merkitysperheen perusmääritelmää
- voiko perusmääritelmää täydentää muilla käsitteillä ja näin selittää tekstissä esiintyviä käsitteen käytössä ilmeneviä merkityseroja.

Toisinaan nuorilla käsitteillä on perheyhtäläinen rakenne: yksi termi nimeää useaa lähikäsitettä tai hierarkkisia käsitteitä. Tällöin tulee selvittää

- mitkä ovat välttämättömät ja riittävät ehdot käsitteen käytölle

– mikä on käsitteen ekstensio
– sopiiko yksi käsitteen käytön normatiivinen rekonstruktio kaikkiin aineiston esiintymiin.

Perusmääritelmän ja ekstension alaa verrataan muihin käsitejärjestelmän käsitteisiin ja yleiseen rakenteeseen ja todetaan, esiintyykö ristiriitaisuuksia tai päällekkäisyyksiä. Lähikäsitteitä, rajatapauksia ja vastaesimerkkejä tutkimalla voidaan tarkentaa, mitä tutkittava käsite ei ole. Koska lähikäsite muistuttaa analysoitavaa käsitettä merkitysrakenteeltaan, erottavat tekijät ovat usein merkittäviä kokonaisteorian näkökulmasta ja siksi myös osa merkitysmatriisia. Merkitysmatriisi osoittaa, miten analysoitava käsite sijoittuu suhteessa ympäröiviin käsitteisiin.

Lopuksi tutkitaan argumentaatioanalyysin avulla, onko käsitteen rekonstruktiolla loogisia implikaatioita tai taustaoletuksia, kun sen sijoittaa takaisin tekstiyhteyksiinsä. Mikäli rekonstruoitu käsite aiheuttaa ristiriitaisuuksia tekstiin, tunnistetaan ja eksplikoidaan ristiriita. Käsitteen analyysi ja tulkinta on riittävä, kun rekonstruoitu käsite, teksti ja käsitejärjestelmän eri osat muodostavat ilmeisen yhtenäisen kokonaisuuden.

# 6. Yhteenveto

Käsiteanalyysi on metodina osa kaikkea tieteellistä tutkimusta, sillä tieteessä täsmennämme käyttämiämme käsitteitä. Metodilla on kuitenkin erityinen rooli kielitieteessä silloin, kun tutkimuksessa on tarve jäsentää jonkin käsitteen täsmällinen merkityssisältö tekstissä. Se, mitä käsiteanalyysin tuloksen viime kädessä uskotaan kuvaavan, riippuu vuorostaan taustalla olevasta merkitysteoriasta. Joidenkin teorioiden mukaan käsitteet ovat mielessä, toisten näkemysten mukaan ne konstruoidaan tekstikontekstissa ja kolmansien mukaan niillä on käsitteen sisäiseen rakenteeseen perustuva itsenäinen viittaussuhde maailmaan. Konkreettisen tekstianalyyttisen käsiteselvittelyn rinnalla on siis tärkeää hahmottaa myös käsitteen metateoreettinen asemoituminen johonkin tieteelliseen taustaan.

**Aiheesta lisää:**

Daly, Christopher (toim.). 2015. *The Palgrave Handbook of Philosophical Methods*. Basingstoke: Palgrave Macmillan.

## VIITTEET

1 Käsitteet on ilmaistu kapiteelein.

## KIRJALLISUUS


Aristoteles. 1994. *Kategoriat. Tulkinnasta. Ensimmäinen analytiikka. Toinen analytiikka*. Jyväskylä: Gummerus.

Dodge, Ellen & Lakoff, George. 2005. Image schemas: From linguistic analysis to neural grounding. Julkaisussa: Hampe, Beate (toim.) *From Perception to Meaning: Image Schemas in Cognitive Linguistics*. Berlin: Mouton de Gruyter, 57–91.

Fodor, Jerry A. 1981. *Representations*. Cambridge (MA): MIT Press.

Forrai, Gabor. 2001. *Reference, Truth and Conceptual Schemes: a Defense of Internal Realism*. Dordrecht: Kluwer Academic Publishers.

Geeraerts, Dirk. 2006. Prototype theory. Julkaisussa: Geeraerts, Dirk (toim.) *Cognitive linguistics: Basic readings*. Berlin: Mouton de Gruyter, 141–165.

Haarala, Risto. 1981. *Sanastotyön opas*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 16. Helsinki: Kotimaisten kielten tutkimuskeskus ja Valtion painatuskeskus.

Hamunen, Markus, Nieminen, Tommi, Kelomäki, Tapani & Dufva, Hannele (toim.). 2019. *Käänteitä ja käsitteitä. Näkökulmia kielitieteelliseen keskusteluun*. Suomalaisen Kirjallisuuden Seuran Toimituksia 1454. Helsinki: Suomalaisen Kirjallisuuden Seura.

Hyvärinen, Matti, Kurunmäki, Jussi, Palonen, Kari, Pulkkinen, Tuija & Stenius, Henrik (toim.). 2003. *Käsitteet liikkeessä. Suomen poliittisen kulttuurin käsitehistoria*. Tampere: Vastapaino.

Jouhki, Jukka & Pennanen, Henna-Riikka (toim.). 2016. *Länsi: käsite, kertomus ja maailmankuva*. Historiallisia tutkimuksia 272. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kakkuri-Knuuttila, Marja-Liisa. 1998. *Argumentti ja kritiikki. Lukemisen, keskustelun ja vakuuttamisen taidot*. Helsinki: Gaudeamus.

Katz, Jerrold & Fodor, Jerry. 1963. The Structure of a Semantic Theory. *Language* 39, 170–210.

Kenttä, Reetta-Leena. 2001. Langackerin kognitiivisen kieliopin metateoreettinen analyysi: Kognitio-käsite ja tieteellinen argumentaatio kognitiivisessa kielitieteessä. Yleisen kielitieteen pro gradu -tutkielma. Helsinki: Helsingin yliopisto.

Konstenius, Reetta A. 2014. *Empiria, eksperimentti ja etiikka: Kielitieteen metateoriaa*. Helsinki: Helsingin yliopisto.
Lakoff, George. 1987. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. Chicago (IL): The University of Chicago Press.
Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar – Volume I: Theoretical Prerequisites*. Stanford (CA): Stanford University Press.
LeCun, Yann & Bengio, Yoschua. 1995. Convolutional networks for images, speech, and time series. Julkaisussa: Arbib, Michael A. (toim.) *The Handbook of Brain Theory and Neural Networks*. Cambridge (MA): MIT Press, 255–258.
Lukose, Dickson, Delugach, Harry, Keeler, Mary, Searle, Leroy & Sowa, John (toim.). 1997. *Conceptual Structures: Fulfilling Peirce's Dream: Fifth International Conference on Conceptual Structures, ICCS'97, Seattle, Washington, USA, August 3-8, 1997. Proceedings* (Vol. 5). Berlin: Springer.
Niiniluoto, Ilkka. 1997. *Johdatus tieteenfilosofiaan. Käsitteen- ja teorianmuodostus*. Helsinki: Otava.
Nirvi, Ruben E. 1952. *Synonyymitutkimuksia sukulaisnimistön alalta*. Suomi 106:1. Helsinki: Suomalaisen Kirjallisuuden Seura.
Rescher, Nicholas. 2001. *Philosophical Reasoning: A Study in the Methodology of Philosophizing*. Malden (MA): Blackwell Publishers.
Riemer, Nick (toim.). 2015. *The Routledge Handbook of Semantics*. New York (NY): Routledge.
Rosch, Eleanor H. 1973. On the internal structure of perceptual and semantic categories. Julkaisussa: Moore, Timothy E. (toim.) *Cognitive Development and Acquisition of Language*. New York (NY): Academic Press, 111–144.
Searle, John R. 1995. *The Construction of Social Reality*. New York (NY): Simon and Schuster.
Taylor, John R. 2008. Prototypes in cognitive linguistics. Julkaisussa: *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York (NY): Routledge, 49–75.
Tegmark, Max. 2014. *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality*. New York (NY): Vintage Books.
Wittgenstein, Ludwig. 1953. *Philosophical Investigations.* Oxford: Blackwell.

# Luku 9 Korpusaineistot

*Veronika Laippala*
https://orcid.org/0000-0002-7635-429X

*Minna Palander-Collin*
https://orcid.org/0000-0001-8428-4423

**Mikä?**

*Korpuksella* tarkoitetaan kielentutkimuksessa laajaa sähköisessä muodossa olevaa tietokoneluettavaa tekstikokoelmaa, joka on strukturoitu ja edustava tutkimuksen tarpeita palvelevalla tavalla. Korpus voi sisältää erilaisia hyödyllisiä lisätietoja, kuten sanaluokkakoodeja tai lauseenjäsennyksen. Kukin korpuksen tekstikatkelma alkaa yleensä metatiedoilla, joista selviää esimerkiksi tekstin kirjoittaja tai puhujat, tekstilaji ja tekstin tuottamisen ajankohta. Näiden merkintöjen avulla korpuksesta voidaan tehokkaasti tehdä erilaisia automatisoituja hakuja.

**Katso myös:**

Luku 7 Laadullinen aineistopohjainen kielentutkimus

Luku 10 Määrällinen korpuslingvistiikka

Luku 11 Historiallinen korpuslingvistiikka

# 1. Johdanto

Korpukset ovat nykyaikaisessa kielentutkimuksessa keskeinen aineistotyyppi, sillä ne tarjoavat tutkijalle empiiristä aineistoa kielenkäytöstä. Niiden avulla voidaan vastata monenlaisiin tutkimuskysymyksiin, testata kielenkäyttöä koskevia hypoteeseja ja tarkistaa esimerkiksi, kuinka tiettyä sanaa käytetään. Tyypillisesti korpuksista saadaan tavalla tai toisella kielestä määrällistä tietoa, jollaista kielen käyttäjien intuitio ei tavoita (ks. Määrällinen korpuslingvistiikka tässä kirjassa). Korpukset soveltuvat kuitenkin myös laadullisten tutkimusten aineistoksi (ks. Laadullinen aineistopohjainen kielentutkimus tk.).

Korpukset ovat tutkijan kannalta nopea keino saada tutkimuksessa tarvittava aineisto, sillä monista kielistä on jo olemassa eri tarkoituksiin koottuja valmiita korpuksia. Oman korpuksen kokoaminenkin voi tuntua helpolta ja houkuttelevalta vaihtoehdolta erityisesti, jos aineisto on jo valmiiksi sähköisessä muodossa esimerkiksi netissä. Korpuksen kokoaminen on kuitenkin aina muutakin kuin vain tekninen ”imurointisuoritus”, ja varsinkin aloittelevan korpustutkijan kannattaa ensin selvittää, millaisia korpuksia on valmiiksi saatavilla. Keskeistä on miettiä, mihin kysymyksiin korpuksella pyritään vastaamaan ja mitä edellytyksiä tämä puolestaan asettaa aineistolle.

Tässä artikkelissa valotamme ensin luvussa 2 lyhyesti korpusten historiaa ja korpustutkimuksen taustalla olevaa kielikäsitystä ja muita tekijöitä. Luvussa 3 esittelemme valmiina saatavilla olevia aineistoja ja luvussa 4 taas tarkastelemme korpusten koostamisessa huomioon otettavia seikkoja. Luvussa 5 esittelemme Kielipankin Korp-käyttöliittymän avulla, miten korpustyökalut toimivat. Lopuksi pohdimme vielä korpustutkimuksen viimeaikaista kehitystä, ns. big datan synnyttämiä uusia metodologisia kysymyksiä ja korpustutkimuksen rajoitteita ja sudenkuoppia.

## 2. Korpustutkimuksen taustaa

Vaikka periaatteessa mikä tahansa systemaattisesti koottu tekstikokoelma voi muodostaa korpuksen, korpuksella tarkoitetaan yleensä nimenomaan tietokoneluettavaa tekstiaineistoa. Korpuslingvistiikan synty nivoutuukin yhteen tietotekniikan kehityksen kanssa. Tärkeä merkitys on ollut myös kielitieteen kehityksellä, sillä viimeistään 1980-luvulta alkaen on tukeuduttu yhä enenevässä määrin aineistopohjaisiin, empiirisiin menetelmiin muun muassa generatiivisen kielentutkimuksen lähes yksinomaisesti käyttämän introspektion eli kielitajun lisäksi (ks. Kielitaju kielentutkijan työkaluna tk.).

The Brown Corpus of Contemporary American English,[1] joka julkaistiin vuonna 1964 Brownin yliopistossa Yhdysvalloissa, oli ensimmäisiä tietokoneluettavia korpuksia. Se sisältää miljoona sanaa vuonna 1961 kirjoitettua amerikanenglantia ja koostuu 500:sta eri tekstikategorioita edustavasta 2 000 sanan otoksesta. Nykymittapuulla Brown-korpus on varsin pieni, mutta sitä käytetään tutkimuksessa edelleen.

Hyödylliseksi Brown-korpuksen tekee se, että muita korpuksia on koottu vastaavalla sapluunalla, kuten vuoden 1961 brittienglantia edustava Lancaster-Oslo-Bergen (LOB) Corpus[2]. Lisäksi alkuperäisten Brown- ja LOB-korpusten verrokit, 1990-luvun alun englantia edustavat Frown[3] ja F-LOB[4] sekä 1930-luvun englantia edustavat B-LOB-1931[5] ja B-Brown[6], mahdollistavat sekä ajallisen että maantieteellisen vertailun. Tällainen aineistojen yhteismitallisuuden ja vertailtavuuden lisääminen voisi yleisemminkin olla paikallaan tutkimuksen läpinäkyvyyden ja toistettavuuden sekä tiedon kumuloitumisen kannalta.

Myös Suomessa korpustutkimus lähti vauhtiin aikaisessa vaiheessa. Oulun yliopistossa laadittiin jo 1960-luvulla suomen kielen Oulun korpus[7], jonka perusteella Pauli Saukkonen (1977) laati muun muassa suomen kielen sanojen yleisyystilaston. Turun yliopistossa sähköisiä teksti- ja ääniaineistoja tarjoava Lauseopin arkisto[8] aloitti toimintansa niin ikään 1960-luvulla. Se sisältää sekä puhutun että kirjoitetun kielen aineistoja, kuten kaikki Suomen murrealueet kattavan murrekorpuksen, arkikeskusteluja sisältävän Arkisyn-korpuksen, Mikael Agricolan

teoksia, vieraskielisten kirjoittamaa suomea käsittävän Edistyneiden suomenoppijoiden korpuksen ja akateemiseen suomeen keskittyvän Akateeminen suomi -korpuksen. Monet näistä on morfosyntaktisesti koodattu niin, että korpuksesta tehtävät haut voidaan kohdistaa esimerkiksi kieliopillisena subjektina toimiviin sanoihin. Helsingin yliopistossa oivallettiin kehittyvän tietotekniikan mahdollisuudet kielentutkimuksessa 1980-luvulla, jolloin koottiin noin 1,5 miljoonaa sanaa historiallista englantia sisältävä The Helsinki Corpus of English Texts,[9] joka Brownin ja LOBin tavoin sisältää näytteitä eri tekstilajeista ja joka ulottuu muinaisenglannin kaudelta varhaisuusenglantiin eli noin vuodesta 730 vuoteen 1710. Helsinki Corpus oli ensimmäinen historiallisen englannin kattava korpus, ja alan tutkijat käyttävät sitä edelleen.

Alkuaikojen korpustutkimus liittyi erityisesti variationistiseen lähestymistapaan, jossa pyrittiin selvittämään, kuinka kielen vaihtelua ja muutosta voidaan ymmärtää systemaattisesti ja kuinka näennäisesti sattumanvaraisia kielellisiä valintoja voidaankin selittää erilaisten kielenulkoisten ja -sisäisten muuttujien avulla (ns. *orderly heterogeineity*; Weinreich, Labov & Herzog 1969). Tällaiseen tutkimukseen korpus on välttämätön työkalu, koska se tarjoaa riittävän määrän systemaattisesti järjesteltyä aineistoa, josta voidaan helposti hakea ja laskea kielen piirteiden frekvenssejä. Korpustutkimus yhdistyy edelleenkin vahvasti määrällisiin menetelmiin ja pyrkimykseen selvittää kielen systematiikkaa, mutta nykyisin myös monet tavallisesti laadullisia menetelmiä soveltavat kielentutkimuksen suuntaukset, kuten diskurssintutkimus ja pragmatiikka, hyödyntävät korpuksia.

Viime vuosikymmenen trendejä korpusrintamalla onkin ollut juuri korpusten käytön leviäminen erilaisten tutkimusperinteiden osaksi (ks. Käännöstiede ja sen menetelmät tk.). Lisäksi erilaiset erikoiskorpukset, puhutun kielen korpukset ja aina vain isommat korpukset ovat lisääntyneet, ja myös korpustyökalut, kuten klusteri- ja avainsana-analyysi sekä muut määrälliset menetelmät, ovat kehittyneet tietojenkäsittelytieteen ja korpuslingvistiikan tutkijoiden yhteistyön myötä (ks. Määrällinen korpuslingvistiikka; Historiallinen korpuslingvistiikka tk.).

Suomessa korpusten käyttöä on tukenut ja edistänyt FIN-CLARINin perustaminen 2000-luvulla. FIN-CLARIN on suomalaisten yliopistojen,

CSC:n ja Kotimaisten kielten keskuksen eli Kotuksen muodostama konsortio, jonka tehtävänä on auttaa esimerkiksi kielentutkijoita käyttämään, säilyttämään ja jakamaan aineistoja. FIN-CLARIN ylläpitää Kielipankkia,[10] jonka kautta tutkijat voivat käyttää laajaa valikoimaa valmiita aineistoja ja myös esittää omia aineistojaan liitettäviksi kokoelmiin muiden käytettäviksi.

Tulevaisuuden korpuksissa voisi ajatella eri modaliteettien, kuten tekstin, äänen ja kuvan, sekä metadatan olevan tutkijan helposti saatavilla siten, että niistä voidaan tehdä ristikkäisiä hakuja ja erilaisia visualisointeja esimerkiksi kaavioiden, karttojen tai animaatioiden muodossa. Näin tulevaisuuden korpustyökalu voisi esimerkiksi havainnollistaa automaattisesti vaikkapa tietyn ajankohtaisen sanan, kuten *ihqu* tai *maahanmuutto,* yleistymistä ajan myötä tai sijoittaa kartalle aineistossa havaittujen murresanojen käyttöä.

## 3. Erilaisia korpuksia

1960-luvun alun jälkeen korpusten koostamisesta on tullut yhä helpompaa ja yleisempää, ja nykyään erilaisia aineistoja on valtavasti saatavilla. Näitä aineistoja voidaan ryhmitellä eri tavoin niiden kokoamisperiaatteiden, käyttötarkoituksen ja niiden sisältämien metatietojen mukaan.

Metatiedot ovat korpukseen konelukuisessa muodossa lisättyjä merkintöjä esimerkiksi korpuksen tekstien kirjoittajista, kirjoitusajankohdasta tai -paikasta tai vaikka puheen prosodiasta. Metatietojen lisäämisestä puhutaan vaihtelevasti joko koodauksena tai annotointina. Tavallinen esimerkki kieliaineistossa on morfologia- tai syntaksiannotoinnit tai -koodaukset, mutta periaatteessa tekstiin voidaan annotoida mitä tahansa tietoja, jotka katsotaan hyödyllisiksi (ks. Määrällinen korpuslingvistiikka; Laadullinen aineistopohjainen kielentutkimus tk.). Oleellista on, että annotoinnit ovat koneellisesti luettavassa muodossa. Näin niitä voidaan käyttää hyväksi koneellisessa analyysissa sekä korpushakuja tehtäessä.

Metatietoja voidaan lisätä joko käsin tai koneellisesti jonkin ohjelman avulla riippuen siitä, mitä metatietoja korpukseen halutaan ja kuinka luotettavia niiden pitää olla. Esimerkiksi korpuksen sanaluokat voidaan monissa tapauksissa tunnistaa automaattisesti riittävällä tarkkuudella. Erityisesti yleiskielelle sanaluokkien tunnistus toimii varsin hyvin, vaikka puheenomaista kieltä tai murteita sisältävien tekstien sanaluokkien tunnistaminen vaatii vielä kehittämistä. Aineistot vaihtelevat sen mukaan, kuinka paljon metatietoa ne sisältävät. Toiset aineistot voivat sisältää hyvin yksityiskohtaiset merkinnät, kun taas toisissa ei ole juurikaan lisättyä tietoa.

Esimerkkinä metatietojen koodaamisesta niin sisällölliseltä kuin tekniseltäkin kannalta voi tutustua avoimesti verkossa saatavilla olevaan Anni Sairion kokoamaan 1700-luvun yksityiskirjeenvaihtoa sisältävään Bluestocking Corpukseen[11], jonka metatiedot löytyvät sekä erillisestä excel-tiedostosta että xml-koodattuina tekstitiedostoista.

Ryhmittelemme alla aineistoja erilaisiin korpustyyppeihin, jotta lukija voi ryhtyä pohtimaan, miten olemassa olevat aineistot voisivat palvella hänen tutkimuksensa tarpeita.

## 3.1. Yleiskorpukset

Näiden korpusten tavoitteena on kuvata kieltä yleisellä tasolla, ja tavallisesti ne pyritäänkin koostamaan eri lähteistä niin, että ne olisivat mahdollisimman kattavia. Monet näistä korpuksista ovat **kansalliskorpuksia** (engl. *national corpora*). Tämä tarkoittaa, että ne kuvaavat jonkin maan kieltä, kuten vaikka British National Corpus[12] englantia tai Venäjän kansalliskorpus[13] venäjää.

## 3.2. Synkroniset aineistot

Synkronisia aineistoja käytetään tavallisesti kuvaamaan kielessä ilmenevää vaihtelua jonakin tiettynä ajankohtana. Näiden avulla voidaan vertailla, miten kieltä käytetään esimerkiksi eri alueilla. Tyypillinen esimerkki on The International Corpus of English (ICE)[14], joka on suunniteltu

eri englannin varieteettien tutkimiseen. Se sisältää 20 miljoonan sanan alakorpusta, joista jokainen edustaa jotain englannin maantieteellistä varieteettia, kuten Australia, Hong Kong tai USA. Vertailun helpottamiseksi jokainen alakorpus on koostettu samoilla kriteereillä, ja niissä on samankaltaiset kielioppimerkinnät.

## 3.3. Diakroniset aineistot

Diakronisia aineistoja käytetään kielen vaihtelun ajalliseen tarkasteluun, eli niistä tutkitaan, miten kieli on muuttunut jonkin ajan kuluessa. Esimerkki diakronisesta korpuksesta on edellä mainittu The Helsinki Corpus of English Texts, joka kattaa noin tuhat vuotta englantia. Diakronisista aineistoista kerrotaan enemmän tämän teoksen luvussa Historiallinen korpuslingvistiikka.

## 3.4. Puhutun kielen aineistot

Sen lisäksi, että yleiskorpuksissa on omat osionsa puhutulle kielelle, on myös aineistoja, jotka sisältävät ainoastaan puhuttua kieltä. The London-Lund Corpus (LLC)[15] on ensimmäinen elektroninen spontaanin puheen korpus. Se sisältää 500 000 sanaa puhuttua brittienglantia. Aineisto on kirjoitettu puhtaaksi eli litteroitu, ja siinä on mukana varsin yksityiskohtaiset merkinnät tekstin prosodiasta.

Myös Suomessa on kerätty monia puhutun kielen aineistoja. Esimerkiksi yllä jo mainittiin arkikeskusteluja sisältävä Arkisyn ja Lauseopin arkiston murrekorpus, jotka molemmat on kieliopillisesti koodattu eli niihin on merkitty sanojen ja virkkeiden morfologisia ja syntaktisia ominaisuuksia. Lisäksi Lauseopin arkistolla ja Kotimaisten kielten keskuksella on koottuna äänitteitä, joita ei ole kaikkia litteroitu: Kotuksen Suomen kielen nauhoitearkisto sisältää noin 24 000 tuntia äänitteitä suomen murteista, ja Lauseopin arkiston äänitearkiston aineistoissa on muun muassa murrehaastatteluja ja luentojen, kokousten yms. tilaisuuksien nauhoitteita.

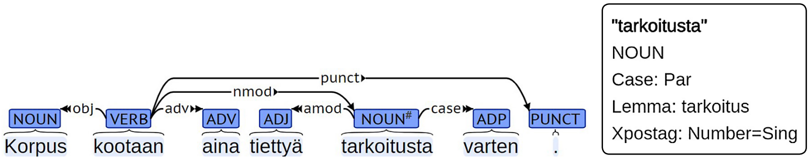
## 3.5. Puupankit eli syntaksiannotoidut korpukset

Kieliaineistoja, jotka sisältävät yleensä käsin tarkistetut syntaksimerkinnät, sanotaan puupankeiksi. Puupankkeja voidaan käyttää aineistoina kielentutkimuksessa. Erityisesti ne soveltuvat tutkimuksiin, joissa tarkastelun kohteena on jokin syntaktinen ilmiö. Syntaksimerkintöjen ansiosta puupankeista voi helposti hakea näitä ilmiöitä. Lisäksi puupankit ovat tarpeellisia kieliteknologiassa. Koneoppimisen avulla niiden pohjalta voidaan kehittää syntaksijäsentimiä eli tietokoneohjelmia, jotka analysoivat kielen morfologiaa ja syntaksia automaattisesti. Näitä voi hyödyntää korpuksen koostamisessa metatietojen lisäämiseen.

Ensimmäinen korpus tässä kategoriassa oli amerikkalainen, vuonna 1992 julkaistu Penn Treebank. Ensimmäinen suomenkielinen vapaasti saatavilla oleva puupankki on Turku Dependency Treebank (TDT), jonka ensimmäinen versio julkaistiin 2009. Nykyään TDT on liitetty kansainväliseen Universal Dependencies -puupankkikokoelmaan, joka sisältää suuren määrän puupankkeja eri kielille.[16]

Varhaisemmissa puupankeissa virkerakenne on usein merkitty lausekerakennekieliopin keinoin (ks. Tieteen termipankki s.v. lausekerakennekielioppi). Nykyään syntaksi kuvataan yleensä dependenssikieliopin avulla (ks. Tieteen termipankki s.v. dependenssikielioppi). Tässä virkkeen rakenne ilmaistaan sanojen välisten riippuvuuksien keinoin. Virkkeen juuri on tavallisesti finiittiverbi, josta lähtee riippuvuudet muihin sanoihin.

Kuviossa 1 finiittiverbi *kootaan* on virkkeen juuri. Juuresta lähtevät objekti (obj), adverbiaalit (nmod ja adv) ja välimerkkiä merkitsevä punct. Sanasta *tarkoitusta* lähtee vielä adjektiivimääriteriippuvuus (amod) tätä määrittävään adjektiiviin *tiettyä* ja case-riippuvuus postpositioon *varten*. Kuviossa sanojen sanaluokka- ja morfologiatiedot ovat näkyvissä vain sanalle *tarkoitusta*, ja niistä käy ilmi, että sanan sanaluokka on NOUN eli substantiivi, sijamuoto partitiivi, lemma eli perusmuoto *tarkoitus* ja luku yksikkö. Todellisuudessa syntaksiannotointien taustalla puupankeissa on vastaavat tiedot kaikille sanoille.




Kuvio 1. Dependenssisyntaksianalyysi virkkeelle.

## 3.6. Internet korpuksena ja koneellisesti kootut aineistot

Internetin kehityksen myötä digitaalisessa muodossa on helposti ja nopeasti saatavilla valtavat määrät tekstiä, josta voi koota kieliaineiston. Ensimmäisiä tällaisia koneellisesti internetistä koottuja aineistoja olivat WaCky Corpora -aineistot[17], jotka sisältävät miljardeja sanoja usealla eri kielellä. Näissä aineistot on koottu rajoittamalla tekstihaku maakohtaisiin päätteisiin, kuten .uk tai .fr. WaCky-aineistojen jälkeen vastaavia aineistoja on koottu yhä laajemmalle kielivalikoimalle. Myös suomelle on Turun yliopistossa koottu, miljardeja sanoja sisältävä Finnish Internet Parsebank, joka on käytettävissä samasta internetkäyttöliittymästä kuin yllä mainittu Universal Dependencies -puupankkikokoelma.[18]

Koneellisesti kootuilla aineistoilla on monia etuja: ne ovat erittäin laajoja, ja ne sisältävät valtavasti kielellistä variaatiota myös tekstilajeista, joita käsin koottuihin aineistoihin harvoin päätyy. Toisaalta niiden käyttöön kieliaineistona liittyy myös haasteita: siinä missä käsin koottujen aineistojen sisältö perustuu huolella pohdittuihin kriteereihin, koneellisesti kootuissa aineistoissa otetaan tavallisesti mukaan kaikki tekstit, jotka sisältävät tietyn vähimmäismäärän haluttua kieltä. Tämän takia koneellisesti kootuista aineistoista on hankala sanoa, millaista kieltä ne sisältävät tai mitä ne oikeastaan edustavat. Tämä pätee myös esimerkiksi valtavaan Google Books -aineistoon, jota voi hakea Google Ngram Viewer -ohjelman avulla.[19]

# 4. Korpuksen kokoaminen

Jos valmiit aineistot eivät sovellu oman tutkimuksen tarpeisiin, tutkija voi päättää koota oman korpuksensa. Korpusta koottaessa on kuitenkin tärkeää muistaa, että korpus ei ole mikä tahansa kokoelma kieltä. Jotta korpus voisi toimia tieteellisen tutkimuksen aineistona, sen on tarkoitus edustaa jotain kieltä tai sen osaa. Tämän takia korpus kootaan aina johonkin tiettyyn tarkoitukseen, ja sen kokoonpanoon vaikuttavat monenlaiset tutkimukselliset ja käytännölliset seikat.

Korpuksen sisältö vaikuttaa tutkimuskysymyksiin, joihin sen avulla voidaan vastata. Esimerkiksi pelkästään radio-ohjelmia sisältävän korpuksen avulla ei voi tarkastella kielen rakenteita puhutussa kielessä ylipäätään, tai pelkästään uutisia käsittävällä aineistolla ei voi vastata kysymyksiin kirjoitetun kielen kehityksestä. Usein puhutaan kriteereistä, joita asiallisesti kerätyn korpuksen tulee täyttää. Esittelemme näitä seuraavissa luvuissa.

Lisäksi korpuksen kokoajan on syytä pohtia ja selvittää jo valmisteluvaiheessa aineiston muotoon, säilytykseen ja jakeluun liittyviä kysymyksiä, sillä nämä eivät ole vain tutkijan harkinnan varaisia seikkoja. Esimerkiksi aineiston säilytyksestä ja jakelusta vastaava infrastruktuuripalvelu voi asettaa aineistolle omia vaatimuksiaan. Aineistoon, sen säilytykseen ja käyttöön voi liittyä myös erilaisia juridisia seikkoja, kuten tekijänoikeus- ja tietosuojakysymyksiä (ks. Johdannon luku 1 tk.). Palaamme näihin lyhyesti tämän luvun lopussa.

## 4.1. Edustavuus

Nykykieltä tutkittaessa on harvinaista, että tutkimuksen kohteena olevaa kieltä tai sen osaa voitaisiin tutkia kokonaisuudessaan. Esimerkiksi, jos tutkimuksen kohteena on jokin tietty rakenne kirjoitetussa nykysuomessa, korpuksena on mahdotonta käyttää kaikkia tekstejä, joita suomeksi on kirjoitettu. Tämän takia tutkimuksessa käytetty korpus on käytännössä aina otos tutkimuksen kohteena olevasta kielestä tai sen variantista.

Siksi kerätyn aineiston pitää edustaa sitä mahdollisimman hyvin. Ainoastaan tällä tavalla tutkimustulokset voidaan yleistää, eli niiden voidaan ajatella koskevan paitsi tutkimusaineistoa myös sen edustamaa kielivarianttia yleensä.

Edustavan otoksen laatiminen on varsin haastavaa, eikä siihen ole yhtä oikeaa vastausta. Tärkeää on sisällyttää korpukseen mahdollisimman laaja valikoima erilaisia kielenkäyttötilanteita siitä kielestä tai sen osasta, jota korpuksen halutaan edustavan. Tavallisesti edustavuuden asettamat haasteet ratkaistaan koostamalla aineisto alakorpuksista, jotka edustavat jotain tiettyä tekstilajia tai kielenkäyttötilannetta. Esimerkiksi yleiskorpus British National Corpus on hierarkkisesti koostettu niin, että aineisto on ensinnäkin jaoteltu puhuttua ja kirjoitettua kieltä käsittäviin osioihin, jotka taas koostuvat erillisistä alaosioista, kuten faktapohjaiset tekstit aiheesta vapaa-aika (engl. *informative: leisure*) ja faktapohjaiset tekstit aiheesta taide (engl. *informative: arts*). Samankaltaisia jaotteluita käytetään myös niin sanotuissa erityisalojen korpuksissa, kuten vaikka oppijoiden kieltä sisältävissä aineistoissa. Esimerkiksi LAS2 – Edistyneiden suomenoppijoiden korpus sisältää tenttivastauksia, tutkielmatekstejä ja esseitä.

## 4.2. Koko

Yleinen pohdinnan aihe korpusta koostettaessa on sen koko. Onko suurempi aina parempi? Mikä on riittävää? Onko ylärajaa? Jos aineisto on liian pieni, siitä saatavat tulokset voivat olla vääristyneitä, mutta isokin aineisto voi antaa kummallisen kapean kuvan kielestä. Iso korpus voi olla sillä tavalla vääristynyt, että se on sisällöltään kovin homogeeninen. Esimerkiksi lähinnä tiettyä genreä, kuten vain kaunokirjallisia tekstejä, sisältävä korpus tuskin kuvaa kovin hyvin kielen kirjoa. Toisaalta valtava aineisto voi olla teknisesti hankalakäyttöisempi kuin pieni ja näppärä korpus, jonka voi tallentaa omalle koneelle ja jota voi selata valmiilla korpusohjelmilla.

Aineiston riittävä koko riippuu myös sen käyttötarkoituksesta. Suuresta koosta on hyötyä esimerkiksi sanastontutkimuksessa, sillä monet

sanat ja kollokaatit eli sanojen yhteisesiintymät voivat olla hyvin harvinaisia. Suuri koko on etu myös yleiskorpuksissa, kuten yllä mainitussa BNC:ssä, joiden tavoitteena on kuvata kieltä yleisesti. Jos yleiskorpus on liian pieni, se ei sisällä riittävästi kielen variaatiota ja kuvaa ainoastaan joitain sen yksittäisiä alaluokkia. Toisaalta tarkempiin tutkimuskysymyksiin tai aihepiirin tai tiheästi esiintyvän kielellisen ilmiön tutkimiseen pienempikin aineisto voi olla riittävä.

Kokoa ei myöskään mitata pelkästään aineiston sanamäärällä. Pitää myös pohtia, otetaanko tekstit aineistoon kokonaisuudessaan vai ainoastaan osina niin, että jokaisessa luokassa on tietty sanamäärä. Lisäksi täytyy ottaa huomioon erilaisten tekstien määrä alaluokissa. Aineiston alaluokkien pitää olla tasapainossa niin, että yksikään luokka ei ole liian vallitseva muihin nähden. Esimerkiksi BNC:ssä kirjoitettuja tekstejä on kokonaisuudesta 90 % ja puhuttuja 10 %. Synkroninen eli kieltä tiettynä ajankohtana kuvaava The Longman/Lancaster Corpus[20] taas sisältää 50 % brittienglantia, 40 % amerikanenglantia ja 10 % muita variantteja, kuten australian-, afrikan- ja irlanninenglantia. Tasapaino täytyy pitää mielessä aina yksilötasolle asti, sillä suhteettoman suuri määrä tekstiä yhdeltä kirjoittajalta tai puhujalta voi tuottaa tuloksia, jotka heijastelevat lähinnä tämän kielenkäyttäjän idiosynkraattisia piirteitä.

## 4.3. Käytännön rajoitteet

Korpuksen koostamiseen liittyy myös monia käytännön rajoitteita, jotka määrittävät lopputulosta eli käytettävää korpusta. Jotkin rajoitteet voivat olla luonnollisia; esimerkiksi historiallisilta aikakausilta ei välttämättä ole saatavilla tiettyjä aineistoja. Toisissa aineistoissa, kuten arkaluonteisia tietoja sisältävissä potilasasiakirja-aineistoissa, saatavuus voi muuten olla rajoitettua. Puhuttuja aineistoja taas voi olla työläs tuottaa.

Ainakin historiallisten kielimuotojen tutkijat ovat kirjoittaneet paljon ”huonon datan” ongelmasta, kun ennen 1900-lukua käytettyjen kielimuotojen tutkimus on perustunut väistämättä ainoastaan kirjallisiin lähteisiin. Koska kielellisten innovaatioiden katsotaan syntyvän epämuodollisessa puheessa, on tätä historiallisessa kielentutkimuksessa

approksimoitu esimerkiksi yksityiskirjeiden ja kirjoitetun puheen, kuten näytelmätekstien, avulla. Lienee kuitenkin parempi käyttää olemassa ja saatavilla olevaa dataa kuin jättää aineistoja kokonaan käyttämättä, vaikka niiden avulla ei voidakaan vastata kaikkiin mahdollisiin tutkimuskysymyksiin.

### 4.4. Avoin saatavuus

Korpuksen koostaminen on työläs prosessi. Lopputulos kuitenkin palkitsee: valmis, huolella koostettu aineisto tarjoaa arvokasta tietoa kielestä ja sen käytöstä. Jotta mahdollisimman monet voisivat hyötyä tästä jo tehdystä työstä, kannattaa varmistaa, että valmis korpus on tallennettu varmaan paikkaan ja että se on mahdollisimman vapaasti saatavilla. Erityisesti 2010-luvulta lähtien aineistojen vapaata saatavuutta onkin korostettu tutkimuksessa. Esimerkiksi Kielipankin kokoelmiin voi luovuttaa valmiin aineiston muiden käytettäväksi (ks. alla). Korpuksen kokoajan on kuitenkin varmistettava, että hänellä on lupa julkaista aineisto avoimesti kenenkään tekijänoikeuksia tai intimiteettiä loukkaamatta. Lupa-asiat ja mahdollinen aineiston metatietojen keruu kannattaa selvittää jo ennen koostamisen aloittamista. Näistä saa lisätietoa esimerkiksi FIN-CLARINin internetsivuilta.[21]

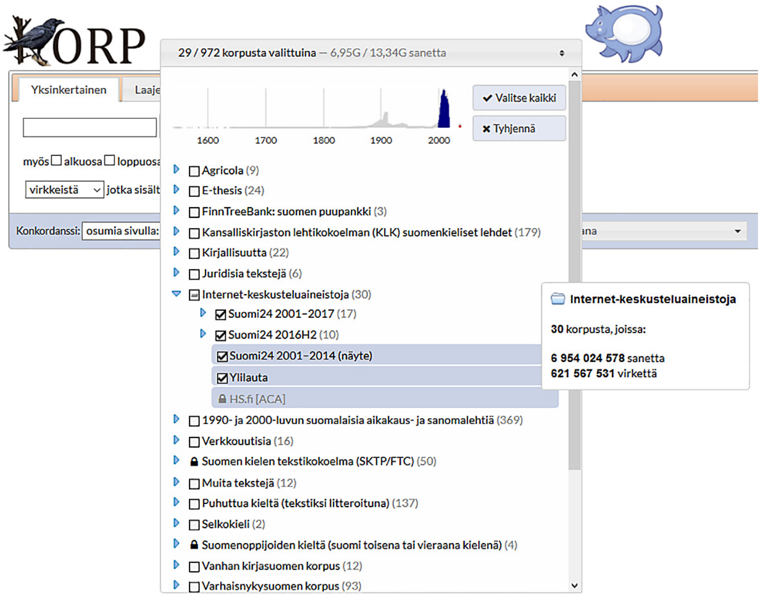
## 5. Korpustyökaluja

Tutkijan johtotähtenä ovat aina tutkimuskysymykset, olipa aineisto mitä tahansa. Annamme tässä joitakin tyypillisiä esimerkkejä siitä, miten korpuksia voi käyttää, mutta teknistä toteutusta tärkeämpää on miettiä, mitä tietoa kyseiset analyysit tuottavat.

Valmis korpus voi olla käytettävissä erilaisissa muodoissa, kuten tekstitiedostoina tai xml-koodattuina tiedostoina, jolloin tarvitaan vielä erillinen ohjelma tai skriptinkirjoitustaitoja korpushakuja varten.

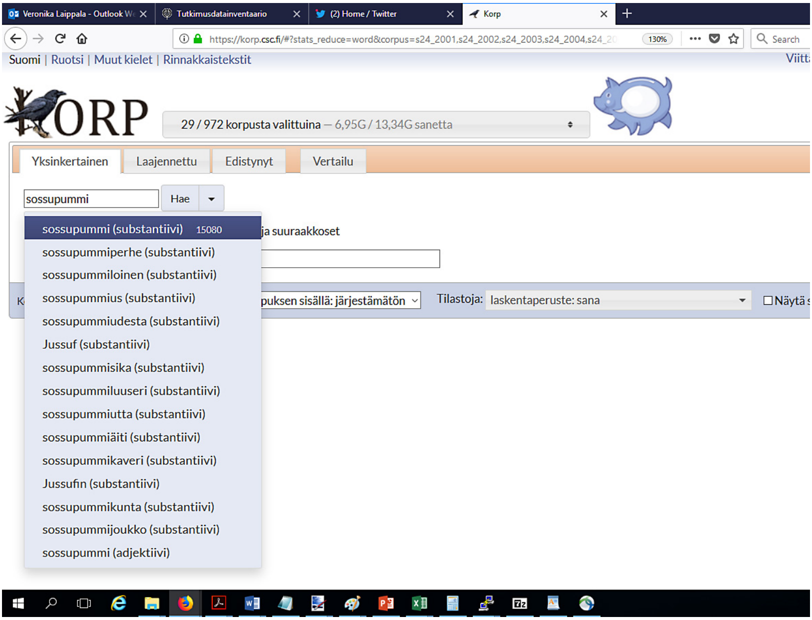
Helpoin tapa useimmille tutkijoille on käyttää korpusta valmiin käyttöliittymän kautta. Esimerkiksi yllä esitellyn Kielipankin aineistoja voi käyttää KORP-hakuliittymän[22] avulla (Borin, Forsberg & Roxendal 2012).

KORP-käyttöliittymä on erittäin monipuolinen ja helppokäyttöinen. Ensimmäiseksi hakua tehdessään käyttäjän tulee valita halutut korpukset monipuolisesta kokoelmasta. Tässä käytämme esimerkkihaussa internetkeskusteluaineistoja, jotka sisältävät yhteensä lähes kolme miljardia sanetta eli sananmuodon esiintymää.



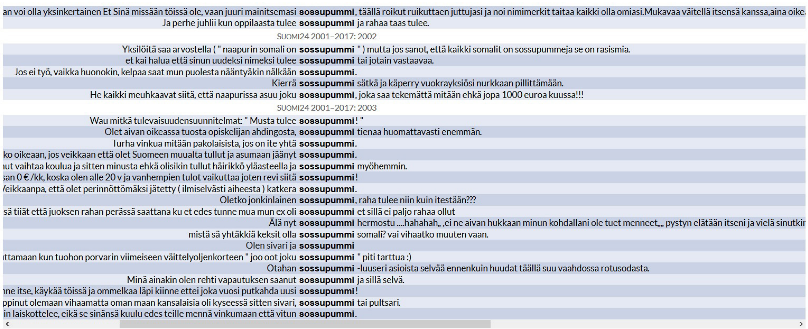

Kuvio 2. Aineiston valitseminen KORPissa.

Aineiston valinnan jälkeen siitä voi tehdä hakuja. Jos aineisto on kielioppillisesti koodattu, haun voi myös kohdistaa tiettyyn sanaluokkaan. Esimerkkihaussa valituissa internetkeskusteluissa näin on, joten ohjelma ehdottaa pudotusvalikossa tarkennuksia hakuun. Haettaessa sanaa *sossupummi* tarkennuksena voi valita esimerkiksi sanaluokan tai jonkin sanan, jonka osana tämä on.




Kuvio 3. Haku KORPissa.

Haun tulokset KORP palauttaa ns. konkordanssinäkymänä. Siinä jokainen hakusana esitetään kontekstissaan omalla rivillään. Hakusanat ovat sijoitettuina allekkain vertaamisen helpottamiseksi.




Kuvio 4. KORPin konkordanssinäkymä.



**Korpus**

Suomi24 (1/10)

Kuvailutiedot
Lisenssi: CC BY-NC (CLARIN PUB)
Viittaa korpukseen
Linkki korpukseen Korpissa: urn:nbn:fi:lb-2015120401

**Tekstin piirteet**

otsikko: Maidon hinnasta 6 senttiä tuottajalle !
otsikon sanojen perusmuodot: maito hinta 6 sentti tuottaja !
päiväys: 09.06.2015
kellonaika: 22:24
keskusteluketjun tunniste: 13638742
viestin tunniste: 79488733
pääaihealue: Yhteiskunta
aihealueen tarkennus: Politiikka > Puolueet > Suomen Keskusta
nimimerkki: kjcfgjj
nimimerkin sanojen perusmuodot: kjcfgjj

**Sanan piirteet**

perusmuoto: sossupummi
perusmuoto (yhdyssanarajat): sossu|pummi
sanaluokka: substantiivi
morfologinen analyysi: NUM_Sg|CASE_Nom
dependenssisuhde: suora objekti
sanan sijainti nimessä: ulkopuolella (O)

alkuperäinen keskustelu
alkuperäinen viesti

Näytä dependenssipuu

Kuvio 5. Tekstin ja haetun sanan kuvailutiedot korpuksessa.

Konkordanssinäkymässä jokaisesta rivistä on myös mahdollista nähdä hakusanan alkuperä. Kuvion 4 esimerkin ensimmäisen rivin tiedot kuvataan kuviossa 5. Siitä käy ilmi, että teksti on peräisin Suomi24-aineistosta (Aller Media Oy 2014). Lisäksi kuvataan tarkemmin tekstin piirteitä ja sanan piirteitä. Näiden avulla hakusanasta saa enemmän ja tarkempaa tietoa.

# 6. Käytä valmiita korpuksia!

Moniin tarkoituksiin sopivia valmiita korpuksia on avoimesti saatavilla, ja erityisesti gradutöitään miettivien opiskelijoiden kannattaa tutustua niihin. Valmis aineisto säästää paljon tutkijan aikaa. Korpuksen metatiedoista voi lukea, millä periaatteilla korpus on koostettu. Aineistona sen pitäisi olla luotettava ja edustava. Sanalla sanoen sen kokoaja on jo ratkaissut monta ongelmaa tutkijan puolesta, ja tutkija voi keskittyä ratkomaan valitsemaansa tutkimuskysymystä. Avoimesti saatavilla oleva aineisto on sikälikin parempi kuin yksittäisen tutkijan tai opiskelijan omaan käyttöön kokoama (pieni) aineisto, että se on myös muiden tutkijoiden käytettävissä, tutkimus on periaatteessa toistettavissa ja sen aukot voidaan osoittaa ja tulosten pohjalta on helpompi suunnitella jatkotutkimuksia.

Vaikka erityisesti jo valmiiksi sähköisessä muodossa olevia tekstejä saattaa olla kohtuullisen helppo koota opinnäytetyön korpukseksi, on kuitenkin korostettava, että huolellisesti kootut laajat korpukset ovat lähes aina ryhmätyön tulosta. Niiden saattaminen valmiiksi ja muiden tutkijoiden saataville kestää vuosia ja vaatii monenlaisia resursseja ja verkostoja, joita yksittäisellä tutkijalla tai opiskelijalla harvemmin on.

Eri kielten tai genrejen osalta valmiiden korpusten saatavuudessa on toki eroja. Englanti on luultavasti maailman tutkituin kieli, joten englannin korpuksia on jo ehditty kokoamaan moneen lähtöön, mikä näkyy tämänkin luvun korpusesimerkeissä, ja lisää esimerkkejä voi hakea Corpus Resource Database (CoRD) -tietokannasta[23]. Kielipankin tarjonta laajentuu jatkuvasti tutkijoiden tallentaessa sinne aineistojaan ja kattaa tällä hetkellä monien suomen kielen korpusten lisäksi useita muita kieliä ja tekstilajeja, kuten Pohjoissaamen korpus, Lettere amorose ja Lyydin korpus.[24] Joistain kielistä valmiita korpuksia ei juurikaan ole, ja olisikin syytä miettiä, voitaisiinko tällaisten kielten tutkimusta edistää kokoamalla avoimesti saatavilla olevia korpuksia.

Valmistakaan aineistoa ei voi käyttää sokkona, vaan sen kokoamisperiaatteisiin on tutustuttava huolella. Ensin kannattaa kaivaa esiin valitun korpuksen kokoamisperiaatteita ja sisältöä selostava käsikirja

eli korpusmanuaali, jos korpuksen kokoajat ovat sellaisen laatineet. Korpusmanuaali on tutkijan keskeinen tietopaketti aineistosta, ja myös oman korpuksen kokoajan on syytä dokumentoida aineistonsa luonne paitsi omaan käyttöön myös mahdollisia muita käyttäjiä varten. ICAME Corpus Manuals -sivustolla[25] voi tutustua erilaisiin korpusmanuaaleihin ja niiden sisältämään tietoon. Lisäksi kannattaa tutustua aiempaan kyseisestä korpuksesta tehtyyn tutkimukseen, jota voi aluksi hakea vaikkapa Google Scholarin avulla, ellei sitä ole koottu yhteen esimerkiksi korpuksen verkkosivulle.

Kävimme edellä läpi joitakin korpuksen kokoamiseen liittyviä yleisiä periaatteita, jotka on hyvä huomioida paitsi oman korpuksen kokoamisessa myös valmiin korpuksen käytössä. Hyvä korpusmanuaali ottaa kantaa juuri edellä kuvattuihin seikkoihin ja esittää kokoamisperiaatteet läpinäkyvästi.

## 7. Korpustutkimuksen rajoitteet ja ongelmat

Ehkä aivan ensimmäinen harha, jonka valtaan korpustutkija saattaa huomaamattaankin joutua, on kuvitella, että korpuksessa on kielen koko kirjo ja ”lopullinen totuus” kielestä. Suomalaisen korpustutkimuksen uranuurtajiin lukeutuva Matti Rissanen (1989) viittaa tähän tutkijan harhaan käsitteellä *God’s Truth Fallacy*. Korpus ei ole koko totuus kielestä, vaan laajoihinkin korpuksiin voi liittyä monenlaisia edustavuuden rajoitteita. Korpukset ovat esimerkiksi paljolti olleet kirjoitetun kielen aineistoja, koska puhetta on paljon vaikeampi kerätä ja muuntaa tekstimuotoon, ja puhekin on saattanut olla tietyntyyppistä suunniteltua tai harkittua puhetta pikemminkin kuin arkipuhetta. Aineisto on usein sitä, mitä saa kohtuullisella vaivalla. On siis täysin mahdollista, että kielellinen ilmiö on olemassa, vaikka sen esiintymiä ei korpuksesta löytyisikään.

Korpuksista ei voi myöskään hakea kaikkea mahdollista automaattisesti, ja haettavuus vaikuttaa siihen, mitä ja miten voidaan tutkia. Sanoja

ja sanan osia voi hakea, ja jos kyseessä on syntaksiannotoitu korpus, kenties sanaluokat tai lauseenjäsenet ovat helposti tutkijan ulottuvilla. Myös semanttisia luokitteluja saattaa sisältyä korpukseen, tai tutkijan on mahdollista luoda niitä itse semanttisten taggereiden eli semanttisia rooleja tunnistavien ohjelmien avulla. Vaikka tällaiset merkinnät helpottavat monella tapaa tutkijan työtä, on kuitenkin tiedostettava, että mitä enemmän hyödyntää valmiita luokitteluja, sitä sidotumpi on tiettyyn kielioppimalliin ja sen rajoitteisiin (ks. Kieliopin tutkimus ja kielioppiteoriat tk.).

Jotkin kielen ilmiöt ovat kerta kaikkiaan vaikeammin haettavia kuin toiset. Niin kauan kuin ilmiön voi sitoa tiettyihin sanoihin tai fraaseihin, hakeminen on helppoa. Pragmaattisia ja vuorovaikutteisia ilmiöitä voi kuitenkin olla vaikea tutkia korpusmenetelmin, jollei niitä pysty operationaalistamaan haettavalla tavalla eli yhdistämällä niitä sanoihin tai fraaseihin.

Korpustutkijat kuvaavat mielellään omaa tutkimustaan sloganilla ”tutkimus alkaa siitä, mihin laskeminen loppuu” (engl. *research begins where counting ends*). Tämä lentävä lause kuvaa hyvin sitä, että laskeminen ei ole korpustutkimuksen itsetarkoitus vaan laskemalla pystytään vastaamaan tutkimuskysymyksiin. Korpustutkimuksen laskennallisuuteen saattaa kuitenkin liittyä myös ongelmia, joista on hyvä olla tietoinen. Kun lingvistiseen koulutukseen ei läheskään aina sisälly edes alkeellista tilastotieteellistä elementtiä, korpustutkijan on erikseen perehdyttävä alansa laskennalliseen puoleen ja kenties värvättävä projekteihinsa alan osaajia. Jotkin korpusohjelmat, kuten WordSmith[26] ja AntConc[27], tarjoavat tilastollisiin menetelmiin perustuvia työkaluja, jolloin tutkijan ei välttämättä tarvitse itse laskea, mutta hänen on kuitenkin pääpiirteissään ymmärrettävä, mitä työkalut hänen puolestaan tekevät.

Korpusten käyttö mahdollistaa kielen tarkastelun erillään kielenkäytön tilanteesta, mutta tutkijan on aina tunnettava korpuksen tekstit ja niiden kielellinen ja kulttuurinen konteksti. Esimerkiksi koneellisesti internetistä koostettuihin aineistoihin voi olla päätynyt osia, jotka vääristävät analyysia, jos kontekstiin ei kiinnitä huomiota: esimerkiksi virke *Lue lisää* esiintyy hyvin usein internetsivuilla muttei silti kerro juuri mitään siellä käytettävästä kielestä. Kielenkäytön kontekstiin tutustuminen voi aineistosta ja tutkimuskysymyksistä riippuen tarkoittaa esimerkiksi korpuksen sisältämien genrejen ominaispiirteisiin ja tuotantoprosesseihin

perehtymistä tai tekstien kirjoittajien henkilöhistorioiden ja keskinäisten verkostojen kartoittamista ja sijoittamista tutkittavan ajan yhteiskunnalliseen kontekstiin. Tässä kontekstualisointityössä tarvitaan yleensä apuna muiden (alojen) tutkijoiden tuottamaa tietoa, joka voi olla luonteeltaan esimerkiksi historiallista tai yhteiskuntatieteellistä.

## 8. Yhteenveto

Tässä artikkelissa olemme käsitelleet korpusten käyttöä ja niiden kokoamista kielentutkimuksessa. Olemme myös esitelleet erilaisia korpuksia ja pohtineet korpustutkimuksen hyötyjä ja mahdollisia sudenkuoppia. Kuten mainitsimme johdannossa, korpusten kokoaminen ja käyttö liittyvät läheisesti yhteen tietotekniikan kehityksen kanssa. Jo nyt kielentutkijoilla on tietokoneiden avulla mahdollisuus päästä käsiksi valtaviin aineistoihin, ja automaattiset ohjelmat helpottavat tutkijan tekemää analyysia esimerkiksi syntaksijäsennyksen avulla. Parhaimmillaan menetelmien kehittyminen synnyttää uudenlaisia tutkimuskysymyksiä ja avaa kieleen sellaisia näkökulmia, joita emme vielä osaa edes ennustaa. Tulevaisuudessa onkin mielenkiintoista päästä seuraamaan, miten ala kehittyy.

**Aiheesta lisää:**

Baker, Paul, Hardie, Andrew & McEnery, Tony. 2006. *Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Biber, Douglas & Reppen, Randi (toim.). 2015. *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press. Saatavissa: https://doi.org/10.1017/CBO9781139764377.

Lüdeling, Anke & Kytö, Merja (toim.). 2008. *Corpus Linguistics. An International Handbook*. Osa 1. Berlin & New York (NY): Mouton de Gruyter. Saatavissa: https://doi.org/10.1515/9783110211429.

Lüdeling, Anke & Kytö, Merja (toim.). 2009. *Corpus Linguistics. An International Handbook*. Osa 2. Berlin & New York (NY): Mouton de Gruyter. Saatavissa: https://doi.org/10.1515/9783110213881.2.

McEnery, Tony & Hardie, Andrew. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press. Saatavissa: https://doi.org/10.1017/CBO9780511981395.

## VIITTEET

1 http://www.helsinki.fi/varieng/CoRD/corpora/BROWN/.
2 http://www.helsinki.fi/varieng/CoRD/corpora/LOB/.
3 http://www.helsinki.fi/varieng/CoRD/corpora/FROWN/index.html.
4 http://www.helsinki.fi/varieng/CoRD/corpora/FLOB/index.html.
5 http://www.helsinki.fi/varieng/CoRD/corpora/BLOB-1931/.
6 http://www.helsinki.fi/varieng/CoRD/corpora/B-BROWN/.
7 https://www.kielipankki.fi/aineistot/oulu/.
8 https://www.utu.fi/fi/yliopisto/humanistinen-tiedekunta/suomen-kieli-ja-suomalais-ugrilainen-kielentutkimus/lauseopin-arkisto.
9 http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/.
10 http://kielipankki.fi/.
11 http://bluestocking.ling.helsinki.fi/.
12 http://www.natcorp.ox.ac.uk.
13 https://ruscorpora.ru/.
14 http://www.ucl.ac.uk/english-usage/projects/ice.htm.
15 http://www.helsinki.fi/varieng/CoRD/corpora/LLC/.
16 https://universaldependencies.org/
17 http://wacky.sslmit.unibo.it/doku.php?id=start.
18 https://universaldependencies.org/
19 https://books.google.com/ngrams.
20 http://www.pearsonlongman.com/dictionaries/corpus/lancaster.html.
21 https://kitwiki.csc.fi/twiki/bin/view/FinCLARIN/FinClarinEsittely.
22 https://korp.csc.fi/.
23 http://www.helsinki.fi/varieng/CoRD/index.html.
24 https://www.kielipankki.fi/aineistot/.
25 http://clu.uni.no/icame/manuals/.
26 http://www.lexically.net/downloads/version6/HTML/index.html?getting_started.htm.
27 http://www.laurenceanthony.net/software/antconc/.

## KIRJALLISUUS


Aller Media Oy. 2014. The Suomi 24 Sentences Corpus (2016H2). [tekstikorpus]. Kielipankki. Saatavissa: http://urn.fi/urn:nbn:fi:lb-2017021505.

Borin, Lars, Forsberg, Markus & Roxendal, Johan. 2012. Korp – the corpus infrastructure of Språkbanken. Saatavissa: https://spraakbanken.gu.se/eng/publikationer/korp-%e2%80%93-corpus-infrastructure-spr%c3%a5kbanken.

Rissanen, Matti. 1989. Three problems connected with the use of diachronic corpora. *ICAME Journal* 13, 16–19.

Saukkonen, Pauli. 1977. *Nykysuomen saneiston yleisyystilastoa saneenloppuisessa aakkosjärjestyksessä*. Oulu: Oulun yliopisto.

Tieteen termipankki. 2019. [verkkoaineisto]. [viitattu 28.2.2019]. Saatavissa: https://tieteentermipankki.fi.

Weinreich, Uriel, Labov, William & Herzog, Martin. 1969. Empirical foundations for a theory of language change. Julkaisussa: Lehmann, W. P. & Malkiel, Yakov (toim.) *Directions for Historical Linguistics*. Austin: University of Texas Press, 97–188.


**Korpusaineistoja**

Esimerkkejä kokoavista metatietokannoista:

- CLARIN Language Resource Inventory: https://www.clarin.eu/content/language-resource-inventory
- Corpus Resource Database (CoRD): http://www.helsinki.fi/varieng/CoRD/
- ICAME Corpus Manuals: http://clu.uni.no/icame/manuals/
- Linguistic Data Consortium: https://www.ldc.upenn.edu
- META-SHARE: http://www.meta-share.org

Keskeisiä monia kieliä sisältäviä korpuskokoelmia:

- CQPweb, Lancasterin yliopiston korpuskokoelma: https://cqpweb.lancs.ac.uk/
- IntraText: http://www.intratext.com/
- Kielipankki: https://www.kielipankki.fi
- Sähköisten kirjojen projekti Gutenberg, ”Project Gutenberg”: http://www.gutenberg.org/
- WaCky-aineistot: http://wacky.sslmit.unibo.it/doku.php?id=start

Keskeisiä suomenkielisiä korpuksia:

- Finnish Internet Parsebank: https://turkunlp.org/finnish_nlp.html#parsebank
- Kansalliskirjaston lehtikokoelma: https://digi.kansalliskirjasto.fi
- Kotimaisten kielten keskuksen sähköiset aineistot: https://www.kotus.fi/aineistot/tietoa_aineistoista/sahkoiset_aineistot_kootusti
- Lauseopin arkisto: https://www.utu.fi/fi/yliopisto/humanistinen-tiedekunta/suomen-kieli-ja-suomalais-ugrilainen-kielentutkimus/lauseopin-arkisto
- Vanhan kirjasuomen korpus: http://kaino.kotus.fi/korpus/vks/meta/vks_coll_rdf.xml

Keskeisiä englanninkielisiä korpuksia:

- British National Corpus: http://www.natcorp.ox.ac.uk/
- Brown Corpus: http://www.helsinki.fi/varieng/CoRD/corpora/BROWN/
  - B-Brown: http://www.helsinki.fi/varieng/CoRD/corpora/B-BROWN/
  - B-LOB-1931: http://www.helsinki.fi/varieng/CoRD/corpora/BLOB-1931/
  - F-LOB : http://www.helsinki.fi/varieng/CoRD/corpora/FLOB/
  - Frown Corpus: http://www.helsinki.fi/varieng/CoRD/corpora/FROWN/
  - Lancaster-Oslo-Bergen (LOB) Corpus: http://www.helsinki.fi/varieng/CoRD/corpora/LOB/
- Corpora of Early English Correspondence (CEEC): http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/
- Corpus of Contemporary American English (COCA): https://www.english-corpora.org/coca/
- Corpus of Historical American English (COHA): https://www.english-corpora.org/coha/
- Corpus of Late Modern English Texts v.3.0 (CLMET3.0): https://perswww.kuleuven.be/~u0044428/clmet3_0.htm
- Early English Books Online (EEBO): https://www.english-corpora.org/eebo/

- ECCO-TCP Corpus: http://www.textcreationpartnership.org
- Helsinki Corpus of English Texts (HC): http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/
- International Corpus of English (ICE): http://www.ice-corpora.uzh.ch/en.html
- The London-Lund Corpus (LLC): http://www.helsinki.fi/varieng/CoRD/corpora/LLC/
- Penn Parsed Corpus of Modern British English (PPCMBE2): http://www.ling.upenn.edu/ppche/ppche-release-2016/PPCMBE2-RELEASE-1

Keskeisiä espanjankielisiä korpuksia:

- ADESSE: http://adesse.uvigo.es/
- Corpus Biblia Medieval: http://corpus.bibliamedieval.es/
- Corpus del español actual: http://cea.spanishfn.org/cea/
- Mark Davies - Corpus del español: http://www.corpusdelespanol.org/x.asp (Käyttääkseen tätä korpusta on rekisteröidyttävä, mutta käyttö on periaatteessa ilmaista.)
- Preseea: http://preseea.linguas.net/Corpus.aspx
- Real Academia Española
    - Corpus del español del siglo XXI (CORPES XXI): http://web.frl.es/CORPES/
    - Corpus de referencia del español actual (CREA) : http://corpus.rae.es/creanet.html
    - Corpus diacrónico del español (CORDE): http://corpus.rae.es/cordenet.html
    - Corpus del Nuevo Diccionario Histórico del Español (CDH): http://web.frl.es/CNDHE/view/inicioExterno.view

Keskeisiä italiankielisiä korpuksia:

- Biblioteca della letteratura italiana: http://www.letteraturaitaliana.net/
- Corpus PAISÀ: http://www.corpusitaliano.it/en/
- Corpus di italiano scritto, CORIS/CODIS: http://corpora.dslo.unibo.it/coris_ita.html

- Listaus italiankielisiin korpuksiin: http://www.accademiadellacrusca.it/it/link-utili/banche-dati-dellitaliano-scritto-parlato

Keskeisiä ranskankielisiä korpuksia:
- Corpus Frantext: https://www.frantext.fr
- Corpora of Computer-Mediated Communication CoMeRe: https://repository.ortolang.fr/api/content/comere/v3.3/comere.html
- Korpuskokoelma Ortolang: http://www.cnrtl.fr/

Keskeisiä ruotsinkielisiä korpuksia:
- Oslon yliopiston pohjoismaalainen murreaineisto: http://www.hf.uio.no/iln/tjenester/kunnskap/sprak/korpus/talesprakskorpus/nordisk-dialekt/index.html
- Suomenruotsalainen murrekorpus TALKO: http://www.sls.fi/talko
- Språkbanken: https://spraakbanken.gu.se/

Keskeisiä saksankielisiä korpuksia:
- Clarin-D-keskusten korpusesittely: https://www.clarin-d.net/en/corpora
- Itävallan Akatemian korpus: http://www.aac.ac.at/
- Katsaus saksankielisiin korpuksiin: Teubert, Wolfgang. 2012. Corpora: German language. Julkaisussa: *The Encyclopedia of Applied Linguistics*. Wiley Online Library, 1–10.
- Puhutun saksan arkisto ”Archiv für gesprochenes Deutsch”: http://agd.ids-mannheim.de/index.shtml
- Saksan kielen referenssikorpus ”Deutsches Referenzkorpus” (DeReKo): http://www1.ids-mannheim.de/kl/projekte/korpora.html
- Saksan kielen digitaalisen sanakirjan (DWDS) korpukset: https://www.dwds.de/
- Sveitsinsaksan korpus ”Schweizer Textkorpus”: https://www.chtk.ch/index.php/de/

- Virheannotoitu oppijansaksan korpus – ”Falko, fehlerannotiertes Lernerkorpus des Deutschen als Fremdsprache”: https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko

Keskeisiä vironkielisiä korpuksia:

- Eesti Keeleressursside Keskus: https://keeleressursid.ee/et/keeleressursid/tekstikorpused.
- Eesti Kirjakeele Korpus 1890–1990: http://www.cl.ut.ee/korpused/baaskorpus
- Eesti keele koondkorpus: https://www.keeletehnoloogia.ee/et/ekktt/ekktt-projektid/eesti-keele-koondkorpus/koondkorpus
- Eesti vahekeele korpus (oppijankielen korpus): http://evkk.tlu.ee
- Tasakaalus korpus: http://www.cl.ut.ee/korpused/grammatikakorpus

Esimerkkejä muunkielisistä korpuksista ja kokoelmista:

- Komi: http://wiki.fu-lab.ru/index.php/%D0%92%D0%B8%D0%BA%D0%B8_FU-Lab
- Portugali: Davies & Ferreira (2006), historiallisen portugalin korpus: https://www.corpusdoportugues.org
- Saame: Tromssan yliopiston saamen kieliteknologia: http://giellatekno.uit.no/index.fin.html
- Tataari: Corpus of Written Tatar: http://www.corpus.tatar/en
  Tatar National Corpus ”Tugan tel”: http://web-corpora.net/TatarCorpus/search/?interface_language=en
- Turkki: Turkish National Corpus: https://www.tnc.org.tr/
- Udmurtti: http://web-corpora.net/UdmurtCorpus/search/index.php?interface_language=en
- Unkari: Unkarin kielen kansalliskorpus: http://corpus.nytud.hu/mnsz/index_eng.html
- Venäjä: Venäjän kansalliskorpus https://ruscorpora.ru/

Esimerkkejä käännöstieteellisen korpustutkimuksen keskeisistä korpuksista:

- Community Interpreting Database (ComInDat): http://www.yorku.ca/comindat/comindat.htm
- The English Norwegian Parallel Corpus (ENPC): http://www.hf.uio.no/ilos/english/services/omc/enpc/
- The European Parliament Interpreting Corpus (EPIC): http://metashare.elda.org/repository/browse/european-parliament-interpretation-corpus-epic/fde33884de7611e2b1e400259011f6ea48ac8ce0b41f48e6be224bbee9d59cb9/
- Käännössuomen korpus. Ks. esim. Mauranen, Anna. 2000. Strange strings in translated language. A study on corpora. Julkaisussa: Olohan, Maeve (toim.) *Intercultural faultlines. Research models in translation studies I. Textual and cognitive aspects*. Manchester: St. Jerome, 119–141.
- The Open Parallel Corpus Opus: http://opus.lingfil.uu.se/
- Translational English Corpus: https://www.alc.manchester.ac.uk/translation-and-intercultural-studies/research/projects/translational-english-corpus-tec/
- YLE-korpus: https://wiivi.uef.fi/crisyp/disp/_/fi/cr_redir_all/fet/fet/sea?direction=1&id=-29931

Esimerkkejä kielitypologian keskeisistä aineistoista:

- AUTOTYP-tietokanta: https://github.com/autotyp/autotyp-data
- The Ethnologue: https://www.ethnologue.com/
- Glottolog: https://glottolog.org
- Typological Database System: https://languagelink.let.uu.nl/tds/index.html
- The World Atlas of Language Structures Online: https://wals.info/
- Cross-Linguistic Linked Data: https://clld.org/
- Database of Places, Language, Culture and Environment: https://d-place.org/

# Luku 10 Määrällinen korpuslingvistiikka

*Aki-Juhani Kyröläinen*

*Veronika Laippala*
https://orcid.org/0000-0002-7635-429X

**Mikä?**

Määrällinen korpuslingvistiikka on kielenkäytön, sen vaihtelun ja kielen rakenteen tutkimusta laajojen aineistojen pohjalta. Analyysi on empiiristä, ja se perustuu laskennallisiin, tietokoneella toteutettuihin menetelmiin. Aineistot ovat tyypillisesti konelukuisia, ja ne on koottu tiettyjen kriteerien mukaan (ks. Korpusaineistot tässä kirjassa).

**Katso myös:**

Luku 7 Laadullinen aineistopohjainen kielentutkimus

Luku 9 Korpusaineistot

Luku 11 Historiallinen korpuslingvistiikka

# 1. Johdanto

Määrällisessä korpustutkimuksessa kieltä tutkitaan laajan, konelukuisen aineiston pohjalta. Määrällisen korpustutkimuksen tavoitteena on, että tutkimusaineiston tulokset eivät kata vain tutkimusaineiston tapauksia vaan ovat yleistettävissä (ks. Korpusaineistot tk.). Esimerkiksi ns. yleiskorpuksiin (engl. *general corpus*) lukeutuvaan British National Corpukseen perustuvien tutkimusten voidaan ajatella kertovan jotakin brittienglannista yleensä, vaikka korpus ei tietenkään sisällä kaikkea brittienglanniksi kirjoitettua tekstiä tai edes kaikkia sen varieteetteja. Tässä artikkelissa lähestymme määrällistä korpustutkimusta juuri tekstiaineistojen näkökulmasta. Havainnollistamme aihetta esimerkkitutkimuksella (Lehti, Luodonpaa-Manni, Jantunen, Kyrölainen, Vesanto & Laippala, tulossa 2020), jossa käytämme aineistona Suomi24-keskustelufoorumin keskusteluja (Aller Media Oy 2014). Keskustelujen voidaan ajatella edustavan laajassa mielessä nykysuomea ja suppeassa mielessä verkkokeskustelupalstoilla käytettävää kieltä (ks. Lagus, Pantzar, Ruckenstein & Ylisiurua 2016).

Määrällisessä korpustutkimuksessa analyysi on empiiristä. Se tarkoittaa, että analyysi ja tutkimuksen tulokset pohjautuvat aineistosta tehtyihin havaintoihin. Tavallisesti määrälliset korpustutkimukset hyödyntävät suuria tekstiaineistoja, joiden katsotaan edustavan tutkimusaihetta. Käytännössä tekstiaineistot perustuvat laajaan otokseen kielestä tai tietyistä kielen ilmiön esiintymistä. Siksi määrällisessä korpustutkimuksessa hyödynnetään laskennallisia menetelmiä, valmiiksi tehtyjä tekstiaineistoja ja tietokoneavusteisia menetelmiä, kuten morfosyntaktisia jäsentimiä. Nämä menetelmät eivät kuitenkaan ole itse tutkimuskohteena, vaan niiden avulla on tarkoitus jäsentää tekstiaineisto analyysia varten.

Usein empiirisiksi kutsutut, määrälliseen aineiston havainnointiin pohjautuvat menetelmät nähdään vastakkaisina intuitioon perustuville menetelmille. Jyrkän vastakkainasettelun sijaan olisi parempi ajatella näitä lähestymistapoja toisiaan täydentävinä näkökulmina kieleen. Lisäksi on syytä muistaa, että vaikka määrälliset, numeerisessa muodossa olevat tulokset antavat pohjan määrälliselle korpustutkimukselle,

laadullinen analyysi on myös olennainen osa sitä. Jotta tutkimustulokset todella merkitsevät jotain, numeeriset tulokset pitää tulkita. Mitä ne tarkoittavat, miten ne vastaavat tutkimuskysymyksiin tai mitä ne kertovat kielestä yleensä? Tämän vuoksi laadullista analyysia ei voi unohtaa myöskään määrällisessä tutkimuksessa.

Määrällinen korpuslingvistiikka voidaan nähdä joko menetelmänä eli lähestymistapana aineistoon tai omana tutkimusparadigmanaan. Tämä jako juontaa juurensa Tognini-Bonellin (2001) esittämään luokitteluun *korpuspohjaiseen* (engl. *corpus-based*) ja *korpuslähtöiseen* (engl. *corpus-driven*) tutkimukseen (ks. Korpusaineistot tk.). Näistä ensimmäinen käyttää aineistoa testaamaan ennalta muodostetun, teoriaan perustuvan hypoteesin, ja tässä korpuslingvistiikka nähdään nimenomaan menetelmänä. Toisessa taas ajatellaan hypoteesien ja tutkimuskohteiden nousevan itse korpuksesta. Tässä korpuslingvistiikka nähdään siis pelkkää menetelmää kokonaisvaltaisempana suuntauksena tai tutkimusparadigmana. Kognitiivisessa kielentutkimuksessa taas todelliseen kielenkäyttöön ja aineistoihin nojautuvaan tutkimukseen viitataan *käyttöpohjaisena tutkimuksena*.

Koska määrällinen korpustutkimus perustuu suuriin aineistoihin ja niistä laskennallisesti saatuihin tuloksiin, sen etuna on, että tulokset ovat yleistettävissä ja kielestä tai tutkitusta yksittäisestä ilmiöstä saa laajemman kuvan. Toisaalta haasteena on, että aineiston kokoaminen on itsessään aikaa vievä tehtävä. Korpuslingvistiikan keinoin on tarkasteltu eniten sellaisia kieliä, joista on olemassa valmiita aineistoja. Erityisesti englanninkielisellä korpuslingvistiikalla on pitkät perinteet, ja englannin kielestä onkin jo olemassa sekä kattavia aineistoja että kehittyneitä tietokoneavusteisia menetelmiä tekstiaineistojen automaattiseen käsittelyyn. Suurten tekstiaineistojen ansiosta yhtäältä tietokoneavusteiset menetelmät ja toisaalta digitaalisten aineistojen kokoaminen ovat kehittyneet ja helpottuneet viime vuosina. Näiden ansiosta tulevaisuudessa voi odottaa, että kielten eriarvoisuus vähenee ja myös vähemmän tutkituille kielille kootaan monipuolisia ja valmiita aineistoja sekä tehokkaita analyysityökaluja.

Useimmiten määrällisessä korpustutkimuksessa tarkastelu kohdistuu sellaiseen kielen ainekseen, jota aineistosta pystytään koneellisesti

laskemaan. Tästä syystä määrällinen korpustutkimus tyypillisesti lähestyy kieltä joko tarkastelemalla kielellisten ilmausten muotojen välistä vaihtelua tai keskittymällä tiettyyn muotoon, esimerkiksi sanaan tai sijamuotoon. Näissä lähtökohtana on muodon suhde merkitykseen eli siihen, mitä merkityksiä se saa ja millaisissa käyttöyhteyksissä sitä käytetään. Toinen vaihtoehto on lähestyä kieltä merkityksestä muotoon -näkökulmasta. Tässä tapauksessa tutkimusaiheena voisi esimerkiksi olla, miten sijaintia ilmaistaan suomen kielessä, ja tutkimus alkaisi sijainnin ilmausten tunnistamisella aineistosta. Tämä on kuitenkin harvinaisempaa, vaikka tietyt merkityslähtöiset tutkimusaiheet voivat näin tehdä. Esimerkiksi juuri nyt pinnalla olevan sentimenttianalyysin eli mielipiteiden automaattisen tunnistamisen tavoitteena on tunnistaa aineistosta kaikki mielipidettä ilmaisevat kielen muodot.

Suomi24-keskustelufoorumin avulla voidaan havainnollistaa kahta edellä mainittua lähestymistapaa. Esimerkkitutkimuksemme tavoitteena on tutkia, miten näissä verkkokeskusteluissa keskustellaan köyhyydestä. Tätä varten olemme eristäneet aineistosta otoksen kommentteja, joissa käytetään sanaa *köyhä* tai sen lähisynonyymisia sanoja, kuten *varaton*. Näiden ilmausten voidaan siis ajatella olevan muotoja, joiden kautta lähestymme yhtäältä *köyhyyden* käsitettä ja toisaalta sen ilmaisemista kielessä. Toisin sanoen tutkimusaihe, köyhyys, on muutettu mitattavaan muotoon lähisynonyymien avulla. Tutkimusaihetta olisi voinut tietysti lähestyä myös merkityksestä muotoon -näkökulmasta. Tämä kuitenkin edellyttäisi, että tutkittava aineisto olisi jo jäsennetty merkityksen ja tutkimuskohteen kannalta tarkoituksenmukaisella informaatiolla. Käytännössä kaikki köyhyyttä käsittelevät kommentit, myös ne, joissa sanaa tai sen lähisynonyymia ei mainita, olisi pitänyt jollain tavalla tunnistaa. Tämä olisi tietysti varsin työlästä. Tämän vuoksi määrällinen korpustutkimus lähestyy kieltä pääsääntöisesti muodosta merkitykseen.

## 2. Historiaa ja tutkimuksen lähtökohdat

Vaikka nykyaikainen määrällinen korpustutkimus perustuu konelukuisiin, digitaalisessa muodossa oleviin aineistoihin ja tietokoneavusteiseen laskentaan, kielenkäyttöä on kuvattu aineistopohjaisesti jo kauan ennen tietokoneita. Jo 1700-luvulla Samuel Johnsonin *Dictionary of English Language* pohjautui kielenkäyttäjien tuottamiin virkkeisiin ja vuonna 1928 julkaistu *Oxford English Dictionary* yli viiteen miljoonaan kielenkäyttäjän sitaattiin, jotka vapaaehtoiset olivat keränneet. Vuonna 1940 julkaistiin Charles Friesin *American English Grammar*, jonka pohjana oli Yhdysvaltojen hallitukselle lähetettyjä kirjeitä.

1960- ja 1970-luvuilla määrällistä korpustutkimusta tehtiin varsin vähän. Silloin vallalla oli generatiiviseen kielitieteeseen perustuva kielikäsitys, jonka mukaan kieli on ennen kaikkea ihmismielen ilmiö ja kielellisen tuotoksen eli aineistojen tutkimuksen ei katsottu olevan tarkoituksenmukaista. Määrällisen korpusanalyysin voidaan kuitenkin katsoa alkaneen 1960-luvun lopulla, jolloin julkaistiin Brown-korpus, yli miljoona sanaa käsittävä elektroninen kieliaineisto (ks. Korpusaineistot tk.). Samankaltaisia aineistoja syntyi tästä lähtien yhä enemmän, ja näitä aineistoja hyödyntäviä, korpuspohjaisia tutkimuksia alettiin julkaista. Suomessa varhaisimpiin korpuksiin lukeutuu Oulun yliopistossa 1960-luvulla koostettu suomen kielen yleiskorpus Oulun korpus[1], jonka perusteella on tehty esimerkiksi suomen kielen sanojen yleisyystilasto (Saukkonen 1977).

Tyypillisesti määrällisen korpuslingvistisen analyysin peruslähtökohtana on laskea tarkasteltavan kielellisen elementin frekvenssi aineistossa, esimerkiksi kuinka usein tiettyjä sanoja tai syntaktisia rakenteita siinä on käytetty. Koska frekvenssillä on keskeinen asema määrällisessä tutkimuksessa, frekvenssiä on tutkittu myös itse ilmiönä (ks. Zipf 1932; Yule 1944). Esimerkiksi korkean frekvenssin sanoilla on tunnettuja ominaisuuksia verrattuna matalan frekvenssin sanoihin. Ensinnäkin ne tyypillisesti muuttuvat hitaammin ajan kuluessa. Tästä ovat esimerkkinä germaanisten kielten epäsäännölliset verbit (Bybee 1995; 2010; Cuskley, Pugliese, Castellano, Colaiori, Loreto & Tria 2014). Toiseksi ne todennäköisemmin vastustavat leksikaalista korvautumista. Esimerkiksi Pagel,

Atkinson ja Meade (2007) arvioivat, että englanin sanan *dirty* 'likainen' leksikaalinen muoto on muuttunut ajan kuluessa yhdeksän kertaa, kun taas sellaiset frekventit sanat kuin *two* 'kaksi' ja *die* 'kuolla' ovat pysyneet samoina. Kolmanneksi frekventit sanat esiintyvät todennäköisemmin yhdessä kulttuurisesti merkittävien tapahtumien kanssa. Esimerkiksi sanan *influenza* 'influenssa' historiallinen frekvenssi osoittaa epidemioiden ajankohdat, kun taas äkillinen muutos sanan frekvenssissä voi heijastaa sensuuria (Michel ja muut 2011). Siten teoreettinen lähtöoletus frekvenssiin perustuvan tutkimuksen taustalla on, että kielen rakenteellisten elementtien, kuten sanojen, vaihtelu heijastaa eroa kielen toiminnassa, kuten sanan merkityksessä tai siinä, millaisissa tilanteissa sanaa käytetään.

Sanan frekvenssi itsessään ei vielä paljasta paljoakaan sanan merkityksestä tai sen mahdollisista käyttöympäristöistä. Siten määrällisen korpustutkimuksen peruslähtökohtana on tarkastella yhteisesiintymiä. Esimerkiksi sanojen *koira* ja *musti* välistä merkityseroa voi lähestyä tarkastelemalla mahdollisia eroja niiden yhteisesiintymissä eli siinä, millaisten sanojen kanssa ne esiintyvät yhdessä. Sana *koira* jakaa kielellistä ympäristöä sellaisten sanojen kanssa kuin *pentu, kissa* ja *koiras,* kun taas *musti* esiintyy samankaltaisissa ympäristöissä kuin *rekku, mirri* ja *hauva.* Näitä yhteisesiintymiä voidaan kutsua myös sanojen *jakaumiksi.* Jo amerikkalainen kielentutkija Harris (1970, 786) totesi, että merkitysero heijastuu erona sanojen jakaumissa.

Sanojen välisillä yhteisesiintymillä on keskeinen asema määrällisessä korpuslingvistiikassa. Esimerkiksi Divjakin ja Griesin (2006) kehittämä käyttäytymisprofiili-menetelmä (engl. *Behavioral profile*) arvioi, millaisten kielellisten aineisten kanssa tietty sana tai rakenne esiintyy yhdessä. Tätä menetelmää on käytetty esimerkiksi venäjänkielisten yrittämistä tarkoittavien verbien (Divjak & Gries 2006) ja englannin *get*-verbin merkitysten tarkasteluun (Berez & Gries 2009). Vastaavaa menetelmää on käyttänyt myös Arppe (2008) tutkiessaan suomen kielen ajattelua merkitseviä verbejä.

Yksittäisten sanojen tai rakenteiden lisäksi voidaan vertailla myös kokonaisia tekstejä. Esimerkiksi Biber (1988) tutki rekistereiden eli tilanteeseen ja päämäärään sidottujen kielen varianttien tyypillisiä piirteitä vertailemalla eri rekistereitä edustavien tekstien sanastoa ja

rakenteellisia ominaisuuksia toisiinsa. Tällöin yhteisesiintymä on yksittäisen tekstin ominaisuuksien (sanasto + rakenne) välillä. Samoin kokonaisten tekstien kohdalla esimerkiksi Biber ja Conrad (2009) toteavat, että rekistereillä on omat frekventit kielelliset piirteensä, koska näitä piirteitä käytetään tekstissä tiettyä tarkoitusta varten; esimerkiksi diskurssipartikkeleita, kuten *hei* ja *joo*, käytetään ilmaisemaan vuorovaikutusta kirjoittajan ja lukijan tai eri kirjoittajien välillä. Vuorovaikutus taas on merkittävä osa esimerkiksi verkkokeskusteluja, minkä vuoksi partikkelit ovat niissä yleisiä.

Yhteisesiintymiä varten on tavallista lisätä aineistoon analyysia syventävää metatietoa. Vaikka esimerkiksi rekistereitä voitaisiin tarkastella sen pohjalta, millainen sanasto on niille tyypillistä, tarkastelun laajentaminen morfologiaan ja syntaksiin tuo rekistereistä enemmän tietoa. Metatietoja voidaan lisätä sekä automaattisesti että käsin. Biberin (esim. 1988) tunnetut tutkimukset rekistereiden ominaispiirteistä hyödyntävät **taggerin** koneellisesti tuottamaa tietoa sanojen merkityksestä ja morfologiasta (ks. taulukko 1). Käyttäytysmisprofiili-menetelmässä sanojen merkitystä tutkitaan aineistoon käsin tehtyjen metatietomerkintöjen perusteella. Merkinnät koskevat esimerkiksi tutkittujen sanojen kanssa yhdessä esiintyvien sanojen morfologisia, syntaktisia ja semanttisia ominaisuuksia. Tässä artikkelissa käyttämämme Suomi24-aineisto taas sisältää keskustelupalstalta automaattisesti tullutta metatietoa muun muassa kommenttien julkaisuajasta ja aihepiiristä.

| Sana | Taggerin analyysi |
|---|---|
| I | pp1a pp1 : *first person subject pronoun + first person pronoun* |
| just | rb emph : *adverbial emphatic* |
| opened | vbd xvbn : *past tense verb + past participle form* |
| my | pp$ pp1 : *possessive determiner + first person pronoun* |
| secret | jj atrb : *adjective + attributive function* |
| bag | nn : *singular common noun* |

Taulukko 1. Taggerin tuottamaa metatietoa (Biber 1988).

Tulevaisuudessa, kun kieliteknologian menetelmät kehittyvät, voidaan koneellisesti tunnistaa ja merkitä yhä monipuolisempia kielen ominaisuuksia. Tällöin käsin tehtävän esikäsittelyn määrä vähenee merkittävästi ja vastaavasti korpuslingvistiikan mahdollisuudet kasvavat entisestään.

## 3. Menetelmiä

Tässä luvussa esittelemme määrällisen korpuslingvistiikan menetelmiä. On tärkeä tuoda esille, että tässä artikkelissa käytetyt ohjelmistot ovat vapaasti saatavilla. Luvuissa 3.1 ja 3.2 käsittelemme perinteisiä ja laajasti käytettyjä menetelmiä, jotka ovat saatavilla erityisesti korpusaineistojen käsittelyyn tarkoitetuissa ohjelmistoissa, kuten AntConcissa ja WordSmithissä. Luvussa 3.3. puolestaan käsittelemme koneoppimista ja sen mahdollisuuksia korpuslingvistiikassa.

Koneoppiminen käsittää luonnollisesti laajan joukon erilaisia menetelmiä. Menetelmät myös kehittyvät koko ajan, eikä niiden käyttö ole vielä vakiintunut korpuslingvistiikassa samalla tavoin kuin luvuissa 3.1. ja 3.2. esiteltyjen menetelmien. Koska koneoppimisella on kuitenkin paljon annettavaa kielentutkimukselle, haluamme tässä artikkelissa valottaa sen toimintaa. Hyödynnämme kahta ohjelmointikieltä, Pythonia ja R:ää. Molemmat ovat laajasti käytössä sekä akateemisessa ympäristössä että yritysmaailmassa. Ne ovat vapaan lähdekoodin ohjelmia ja vapaasti saatavilla. Perinteisesti Python on painottunut enemmän tietojenkäsittelyyn, kun taas R:ää käytetään enemmän tilastotieteessä. Molemmat tarjoavat monipuolisia ja edistyneitä toimintoja aineiston analyysiin ja soveltuvat siksi hyvin määrälliseen korpustutkimukseen.

## 3.1. Avainsana-analyysi ja avaimuus

Avainsana-analyysista on tullut yksi määrällisen korpuslingvistiikan yleisimmistä menetelmistä. Varhaisimmissa tutkimuksissa avainsanoilla viitattiin käsitteisiin, jotka heijastavat kulttuurin tai yhteiskunnan ominaisuuksia. Korpuslingvistiikassa *avainsanoilla* tarkoitetaan kuitenkin tavallisesti sanoja, jotka kuvaavat jollain tavalla merkittäviä tekstin ominaisuuksia. Esimerkiksi Ivaskan (2015, 51) mukaan "termillä avaimuus halutaan korostaa havaittujen erojen merkitystä avaimena sillä saavutettavaan tietoon käsillä olevan aineiston erityispiirteistä".

Avainsanat ovat siis sanoja, jotka jollain tavalla toimivat "avaimena" tekstiin. Menetelmää voidaan käyttää sekä tekstilajien tyypillisten piirteiden määrittämiseen että analysoitujen tekstien välittämien diskurssien[2] kuvaamiseen. Esimerkiksi Baker, Gabrielatos, Khosravinik, Krzyzanowsky, McEnery ja Wodak (2004) analysoivat uutistekstejä ja Roemer ja Wulff (2008) puolestaan opiskelijoiden esseitä. Analyysin tuloksena syntyi lista näitä tekstejä kuvailevia avainsanoja. Ne jaotellaan usein kolmeen luokkaan: 1) erisnimiin, 2) sanoihin, jotka ihmislukijakin tunnistaisi avainsanoiksi ja jotka kertovat tekstin aihepiiristä, ja 3) yleisiin sanoihin, jotka kertovat pikemminkin tekstin tyylistä kuin aihepiiristä.

Avainsana-analyysin lähtökohtana on, että avainsanojen keskeinen rooli ei ole vain sattuman tulosta aineistossa vaan avaimuus edustuu sanoissa, joiden esiintyminen ei ole odotusarvon mukaista. Odotusarvo lasketaan vertaamalla tutkimusaineiston sanojen frekvenssejä laajempaan verrokki- tai referenssiaineistoon. Sanan frekvenssi on odotusarvon mukainen, jos se ei eroa tutkimus- ja verrokkiaineiston välillä. Avainsanat puolestaan esiintyvät tutkimusaineistossa odotusarvoa enemmän tai vähemmän. Frekvenssierojen tilastolliseen todentamiseen käytetään yleensä $\chi 2$-testiä tai log-uskottavuutta (engl. *log likelihood*). Avainsanatoiminto on usein valmiina korpusohjelmissa, kuten AntConcissa tai WordSmithissä (ks. Historiallinen korpuslingvistiikka tk.). Avainsanat esitetään listattuna siten, että merkittävimmät ovat listan alussa. On tärkeä muistaa, että avainsanoissa ovat mukana sekä positiiviset eli aineistossa yliedustetut avainsanat, joita on odotusarvoa enemmän, että negatiiviset eli aineistossa aliedustetut avainsanat, joita on odotusarvoa

vähemmän. Viimeksi mainitut kuvaavat aineistoa kääntäen, eli niiden poissaolo on siinä tyypillistä.

Avainsanojen määrittäminen vertaamalla tutkitun aineiston sanastoa verrokkiaineiston sanastoon on ehdottomasti tavallisin tapa laskea ne. Tästä huolimatta tällä menetelmällä on selviä puutteita. Ensinnäkin se käsittelee koko aineistoa yhtenä kokonaisuutena. Tämä on ongelmallista, jos aineiston tekstit ovat keskenään erilaisia. Esimerkiksi tietyt sanat saattavat olla joissain tutkimusaineiston teksteissä erittäin yleisiä mutta toisissa taas harvinaisia. Kun avainsanat lasketaan yhdistämällä kaikki aineiston tekstit yhdeksi kokonaisuudeksi, nämä erot yksittäisten tekstien välillä eivät näy tuloksissa. Toiseksi sanastojen välisten erojen tarkastelu tilastollisella todennäköisyydellä voi olla harhaanjohtavaa suurissa aineistoissa. Tutkimusaineiston ja verrokkiaineiston suuren koon takia pienetkin erot sanastoissa voivat olla tilastollisesti merkitseviä, eli sanastojen frekvenssierot eivät ole sattuman tulosta. Tilastollinen merkitsevyys ei kuitenkaan tarkoita sitä, että sanastojen väliset erot olisivat automaattisesti merkitseviä tutkimuskysymysten kannalta.

Analyysimenetelmän puutteiden kehittämiseksi on ehdotettu erilaisia ratkaisuja. Esimerkiksi *key keyword analysis* -nimisessä menetelmässä koko aineiston avainsanat kerätään joukosta niitä sanoja, jotka ovat avaimia myös yksittäisissä teksteissä. Näin tekstien väliset erot eivät pääse vääristämään tuloksia. Toisaalta tilastollisten menetelmien parantamisen sijaan voi olla yksinkertaisempaa soveltaa ohjattua koneoppimista avaimuuden tutkimiseen. Tarkastelemme tätä lähestymistapaa tarkemmin luvussa 3.3.

## 3.2. Kollokaatioanalyysi

Avaimuuden lisäksi toinen perinteinen määrällisen korpustutkimuksen menetelmä on kollokaatioanalyysi. Jo korpustutkimuksen varhaisvaiheissa J. R. Firth määritti, että sanan tuntemiseen vaaditaan sen kontekstin tunteminen: ”You shall know a word by the company that it keeps” (Firth 1957, 6). **Kollokaatio** tarkoittaa siis sanojen taipumusta esiintyä yhdessä. Sanaa ja sen kollokaattia eli lähiympäristössä

esiintyvää sanaa käytetään yhdessä odotusarvoa enemmän. Esimerkiksi verbin *naukua* subjektina voisi toimia usein *kissa* tai *pentu*, joskin myös *ovi naukuu* ja *nälkä* voi *naukua vatsassa*. Monet kollokaatit muodostavat pääsanansa eli rakenteen muuttumattoman **ytimen** kanssa vakiintuneita ilmauksia, kuten *rauhanomainen rinnakkaiselo*.

Tavallisesti kollokaatit kertovat jotain ytimen (kuten verbin *naukua*) merkityksestä. Kollokaatioanalyysia käytetäänkin sanojen merkityksen eli semantiikan tutkimukseen. Voidaan ajatella, että merkitys ei rakennu pelkästään yksittäisistä sanoista vaan laajemmista kokonaisuuksista ja että sanoihin liittyy helposti niiden kollokaattien merkitystä. Esimerkiksi Stubbs (1995) on kollokaatioanalyysin avulla osoittanut, että englannin verbi *cause* esiintyy tyypillisesti negatiivisten ilmausten yhteydessä.

Keskeinen käsite kollokaatioanalyysissa on **semanttinen preferenssi**. Tällä tarkoitetaan sanojen taipumusta esiintyä yhdessä keskenään samaan merkityskenttään kuuluvien sanojen kanssa. Esimerkiksi Stubbs (1995) raportoi englannin sanalle *cause* kollokaateiksi muun muassa sanat *abandonment* 'luovuttaminen', *accident* 'onnettomuus', *alarm* 'huoli' ja *anger* 'viha'. Nämä ovat hyviä esimerkkejä negatiivisista ilmauksista, jotka nousevat Stubbsin artikkelissa esiin *cause*-verbin kollokaatteina.

Tietylle ytimelle tai tutkitulle sanalle voidaan myös muodostaa ns. **semanttinen profiili**. Se koostuu korpusaineiston perusteella määritellyistä kollokaateista, jotka ovat ytimelle tyypillisiä. Näiden avulla voidaan paitsi kuvata yksittäisten sanojen merkitystä myös vertailla sanoja ja aineistoja toisiinsa. Esimerkiksi Jantunen (2012) analysoi *tässä*-pronominin kollokaatteja suppeammalle asiantuntijayleisölle tarkoitetuissa akateemisissa teksteissä ja laajalle yleisölle kohdennetuissa populaariteksteissä, kuten elämäntaito-oppaissa. Hän jakaa *tässä*-pronominin kollokaatit kolmeen ryhmään: 1) metatekstuaalisiin ilmaisuihin, joilla viitataan itse tekstissä raportoitavaan tutkimukseen (*tässä kirjassa, tässä luvussa, tässä tutkimuksessa*), 2) näkökulmaa ilmaiseviin yhteisesiintymiin (*tässä perinteessä, tässä kohden*) ja 3) abstraktimpiin, yksittäisiin tilanteisiin viittaaviin konstruktioihin (*tässä tapauksessa, tässä vaiheessa*).

Jantusen tulokset kertovat tässä-pronominin käytöstä ja merkityksestä huomattavasti enemmän kuin pelkät frekvenssit tekisivät. Myös tekstilajien erot näkyvät kollokaattien frekvensseissä jonkin verran: esimerkiksi

metatekstuaaliset ilmaukset ovat akateemisissa teksteissä populaaritekstejä yleisempiä. Tämä kertoo akateemiselle tekstille tyypillisestä tavasta jäsentää tekstiä lukijalle ja rakentaa sen välittämää viestiä viittaamalla tekstin osiin eksplisiittisesti.

Monet yleiset korpusohjelmat, kuten AntConc, sisältävät työkalun, joka määrittää kollokaatit halutuille sanoille. Keskeistä tässä lähestymistavassa on se, että kollokaatioksi muodostuu sellainen sanojen yhteisesiintyminen, joka poikkeaa odotusarvosta. Sanat eivät vain esiinny yhdessä, vaan niiden yhteisesiintyminen poikkeaa siitä, mitä voisimme aineiston perusteella olettaa. Tästä syystä kollokaatioanalyysi on joustava menetelmä ja siinä voidaan hyödyntää useita muita menetelmiä, esimerkiksi **log-uskottavuutta** tai **keskinäistä informaatioarvoa** (engl. *mutual information*) (Anthony 2014). On syytä painottaa, että eri menetelmät eivät välttämättä tuo esille aineistosta samoja piirteitä vaan ne saattavat korostaa tiettyjä osa-alueita. Esimerkiksi keskinäinen informaatioarvo nostaa esille kollokaatteja riippumatta sanojen havaitusta frekvenssistä, kun taas log-uskottavuus ottaa tämän huomioon.

## 3.3. Koneoppiminen korpuslingvistiikassa

*Koneoppimisella* viitataan menetelmiin, jotka hyödyntävät yleisiä algoritmeja, joiden tavoitteena on oppia aineistosta. Koneoppimisalgoritmi ei ole vain staattinen kokoelma käskyjä, jotka olisi luotu tiettyä aineistoa varten, vaan se oppii mallin aineistosta ja ennustaa tulevia toteutumia tämän mallin perusteella. Koneoppiminen jaetaan kahteen päätyyppiin, ohjattuun ja ohjaamattomaan. **Ohjattua koneoppimista** hyödynnetään silloin, kun tavoitteena on oppia vastemuuttuja syötteestä. Esimerkiksi vastemuuttuja voisi olla sähköpostien jaottelu kahteen luokkaan, roskapostiin ja asiallisiin viesteihin. Ohjelman pitäisi ennustaa tämä vastemuuttuja eli viestin luokka sille annetun syötteen eli itse viestin perusteella. **Ohjaamatonta koneoppimista** hyödynnetään taas silloin, kun tavoitteena on löytää aineistosta rakenne ilman etukäteen määriteltyjä luokkia. Tällöin esimerkiksi sähköpostiviestejä voitaisiin ryhmitellä samankaltaisiin ryhmiin. Molemmille koneoppimistyypeille on yhteistä se, että tutkimusaineisto

on ensin muutettava piirteiksi. Tällä tarkoitetaan sitä, miten aineisto esitetään ohjelmalle. Piirteistys on keskeistä koneoppimisen soveltamisessa kielentutkimukseen, sillä se määrää, miten tutkimusaihe muutetaan mitattavaan muotoon eli **operationalisoidaan**. Käsittelemmekin aineiston piirteistystä luvussa 3.3.1. Tämän jälkeen siirrymme ensin ohjattuun ja sitten ohjaamattomaan koneoppimiseen.

### 3.3.1. AINEISTON PIIRTEISTYS ELI MITEN MUUTTAA TUTKIMUSAIHE MITATTAVAAN MUOTOON

Piirteistyksen tarkoituksena on operationalisoida tutkimusaihe eli muuttaa se mitattavaan muotoon. Esimerkkitutkimuksessamme luvussa 4 tarkastelemme sitä, miten köyhyydestä keskustellaan Suomi24-palstalla. Tätä varten ensimmäinen askel on hakea aineistosta kaikki kommentit, joissa mainitaan sana *köyhä* tai jokin sen lähisynonyymi. Toisin sanoen *köyhyyden* operationalisointi aloitetaan näkemällä se kimppuna lekseemejä, jotka liittyvät merkityksen kannalta käsitteeseen *köyhä*. Toinen askel on muuntaa tarkasteltava kielenaines eli keskustelupalstakommentit mitattavaan muotoon ja piirteistää ne.

Käytetyt piirteet voivat olla periaatteessa mitä tahansa. Määrällisessä korpuslingvistiikassa piirteiden valinta on kuitenkin yksi keskeisimmistä haasteista. Haasteellisuus piilee siinä, että toisaalta tutkimuksessa käytettyjen piirteiden avulla algoritmin on opittava tunnistamaan tutkittavat kielen ilmiöt, esimerkiksi roskapostiviestit tai köyhyyttä käsittelevät keskustelupalstakommentit, ja toisaalta piirteiden tulee olla kielitieteellisesti motivoituja. On olennaista, että piirteet kuvaavat tarkasteltavia ilmiöitä ja että niiden avulla saadaan uutta tietoa ilmiöistä. Tavallisesti piirteiden valinta perustuukin ennen kaikkea tutkimusasetelmaan, johon vaikuttavat sekä tutkijan kielitaju (ks. Kielitaju kielentutkijan työkaluna tk.) että aiempi tutkimus. Tiedetään esimerkiksi, että kielen rakenteelliset eli morfosyntaktiset ominaisuudet vaihtelevat eri rekistereitä edustavissa teksteissä. Näin ollen rekistereiden kielellisiä ominaisuuksia tarkastelevassa tutkimuksessa nämä ominaisuudet kannattaa ottaa piirteisiin mukaan. Sen sijaan esimerkiksi se, minä viikonpäivänä teksti on kirjoitettu, ei todennäköisesti kuvaa rekistereiden välisiä mahdollisia eroja, eikä tätä ominaisuutta kannata huomioida piirteistyksessä.

Piirteiden valintaa voi lähestyä kahdesta näkökulmasta, aineisto- tai teoriavetoisesti. Yksi perinteinen aineistovetoinen piirteistys on **sanasäkkimalli** (engl. *bag-of-words*). Tekstiaineiston, joko lauseen tai pidemmän tekstin, kuten dokumentin, esitetään koostuvan ainoastaan kokoelmasta sanoja ja niiden frekvensseistä. Esimerkiksi sanajärjestystä ei huomioida. Sanasäkkimalli voidaan toteuttaa joko sane- tai lemmatasolla. **Saneella** tarkoitetaan sananmuodon esiintymää tekstissä, esimerkiksi *talossa* ja *pihalla*. Termi **lemma** viittaa taas sanan ns. sitaattimuotoon, joka edustaa tietyn sanan kaikkia taivutusmuotoja; esimerkiksi saneiden *taloa*, *talosta* ja *taloon* lemma on *talo*. Sanan lemmaa käytetaan tyypillisesti sanakirjoissa hakusanana.

Sanasäkkimallin toteuttamisvalinnalla on yhtymäkohtia sekä aineistoon että itse tutkimuskysymykseen. Aineiston lemmatisointi vaatii tietokoneavusteisen menetelmän hyödyntämistä. Lemmatisointi voi kuitenkin myös tuoda mukanaan virhettä aineistoon, eli menetelmä voi tuottaa väärän lemmamuodon saneelle. Teknisen toteutuksen lisäksi tutkimuskysymys ohjaa myös sanasäkkimallin toteuttamistapaa. Jos tutkimuskysymys liittyy kiinteästi kysymykseen *mistä on kirjoitettu*, aineiston lemmatisointi on hyvä lähtökohta. Jos taas tutkimuskysymys on pikemminkin *miten on kirjoitettu*, aineiston lemmatisointi voi olla ongelmallista, koska lemmatisointi hävittää morfologian. Lemmatisoinnin merkittävin hyöty on kuitenkin siinä, että se vähentää piirteiden määrää, koska lemmapohjaisessa sanasäkkimallissa piirteiden määrä on sama kuin aineiston uniikkien lemmojen määrä. Sanepohjaisessa mallissa taas piirteiden määrä tulee saneiden mukaan.

Havainnollistamme sanasäkkimallia saneilla taulukossa 2, jossa riveillä on dokumentit ja sarakkeilla saneet ja soluarvot ovat frekvenssejä.

| | talossa | pihalla | talosta | pihalta |
|---|---|---|---|---|
| dokumentti 1 | 40 | 60 | 0 | 0 |
| dokumentti 2 | 0 | 0 | 60 | 80 |
| dokumentti 3 | 0 | 40 | 0 | 60 |
| dokumentti 4 | 120 | 0 | 60 | 0 |

Taulukko 2. Sanasäkkimalli saneiden perusteella.

Frekvenssin absoluuttisen arvon perusteella dokumentit 1 ja 2 eroavat sijamuotojen suhteen, eli näitä dokumentteja erottaa ilmauksen funktio. Dokumentit 3 ja 4 puolestaan eroavat sisällön perusteella: *talo* versus *piha*. Taulukossa 3 on kuvattu sama aineisto lemmatisoituna.

| | talo | piha |
|---|---|---|
| dokumentti 1 | 40 | 60 |
| dokumentti 2 | 60 | 80 |
| dokumentti 3 | 0 | 100 |
| dokumentti 4 | 180 | 0 |

Taulukko 3. Sanasäkkimalli lemmatisoidulla aineistolla.

Lemmatisoidussa aineistossa sisältöperusteiset erot ovat säilyneet (ks. dokumentti 3 ja 4) mutta morfologinen ero on kadonnut. Merkittävin etu lemmatisoidussa aineistossa on kuitenkin se, että piirteitä on huomattavasti vähemmän kuin sanepohjaisessa mallissa. Toisin sanoen aineisto on tiheämpää. Sanepohjaisessa sanasäkkimallissa aineisto taas on harvempaa, eli uniikkeja saneita on enemmän.

Sanasäkkimallin lisäksi piirteistyksessä on mahdollista huomioida myös sanojen peräkkäinen yhteisesiintyminen. Tällaisia malleja kutsutaan yleisnimellä *n-grammimalli*. Bigrammimalli perustuu kahden sanan yhteisesiintymiseen aineistossa ja trigrammi puolestaan kolmen. Siten aineiston esitetään koostuvan tekstissä peräkkäin esiintyvien sanojen yhdistelmistä, joissa sanojen määrä on N. Sanasäkkimalli on siis n-grammimallien erityistapaus, unigrammimalli. On syytä huomioida, että n-grammimallien tavoitteena on tavallisesti juuri aineiston piirteistys koneoppimista varten. Yhteisesiintymiin perustuvan kollokaattianalyysin (ks. luku 3.2.) tavoite puolestaan on tarkastella sanojen merkitystä.

Tavallisesti n-grammimalli ei kuitenkaan huomioi sanojen morfologisia piirteitä tai niiden välisiä syntaktisia suhteita, jotka saattavat yhdistää virkkeessä toisistaan kaukanakin esiintyviä sanoja. Laippala, Kyröläinen, Kanerva ja Ginter (2018) laajentavat juuri n-grammimalleja suomen kielessä ottamalla huomioon sanojen väliset syntaktiset suhteet. Tietysti erilaisia n-grammimalleja on mahdollista myös yhdistää

ja vertailla. Esimerkiksi Laippala, Luotolahti, Kyröläinen, Salakoski ja Ginter (2017) vertaavat, miten erilaiset n-grammimallit ja niiden yhdistelmät soveltuvat suomenkielisessä internetissä käytettävien rekisterei-den tunnistamiseen.

Sanasäkki- ja n-grammimallien kohdalla on syytä kuitenkin muistaa, että aineiston esikäsittelyllä on merkittävä osuus tutkimuksessa. Näin on erityisesti verkkoaineiston kohdalla, sillä se saattaa sisältää paljon analyysin kannalta ei-toivottua ainesta, joka ei lisää informaatiota vaan tuo mukanaan kohinaa. Esimerkiksi virke *Lue lisää* saattaa esiintyä koneellisesti kootussa aineistossa usein sivustojen teknisen toteutuksen takia (ks. myös Korpusaineistot tk.). Sillä ei kuitenkaan ole juurikaan kielentutkimuksellista selitysvoimaa, vaikkapa kielen rekisterien ominaispiirteiden tarkastelussa.

Siinä missä aineistovetoisessa piirteistyksessä esimerkiksi sanasäkkimalli on tarkoitettu yleiseksi malliksi kieliaineiston määrälliselle korpustutkimukselle, teoriavetoisessa piirteistyksessä aineiston piirteet valitaan yksittäin ja piirteistys on tyypillisesti sidottu konkreettiseen tutkimuskysymykseen. Esimerkiksi Helasvuo ja Kyröläinen (2016) mallintavat suomen kielen ensimmäisen persoonan subjektiargumentin reaalistumista keskusteluaineistossa joko pronominina (*minä luen*) tai nollasubjektina (*luen*). Piirteinä käytetään aiemmissa tutkimuksissa havaittuja ja teoreettisesti motivoituja muuttujia. Verbeihin liittyviä piirteitä ovat ensinnäkin sen semanttinen luokittelu esimerkiksi kognitioverbeihin (*ajatella*) ja liikeverbeihin (*juosta*). Toisena piirteenä on verbien frekvenssi, jossa käytetään lemmamuotojen frekvenssiä. Lausetason piirteenä puolestaan voidaan mainita polaarisuus eli lauseen myönteinen tai kielteinen sisältö.

### 3.3.2. OHJATTU KONEOPPIMINEN

Ohjatussa koneoppimisessa oppiminen perustuu valmiisiin esimerkkeihin, jotka ovat usein joko tutkijan itsensä kielitajun pohjalta käsin tekemiä tai peräisin valmiista aineistosta. Näiden valmiiden esimerkkien kokoelmaa kutsutaan *harjoitusaineistoksi*. Vaikkapa roskapostit tunnistavan ohjelman harjoitusaineisto voisi koostua viesteistä, jotka on merkitty roskaposteiksi, ja viesteistä, jotka on merkitty tavallisiksi, asiallisiksi

viesteiksi. Esimerkkiaineiston perusteella algoritmi pyrkii oppimaan luokkien ominaisuudet niin, että se pystyy luokittelemaan uudet viestit näihin kahteen luokkaan. Tässä luvussa keskitymme ohjatun koneoppimisen soveltamiseen nimenomaan luokittelussa. Kielentutkimuksessa ohjattua koneoppimista ja luokittelua käytetään esimerkiksi kielivarianttien tai joidenkin kielen rakenteiden ominaispiirteiden tutkimiseen. Esimerkiksi luvussa 3.3.1. esitelty suomen kielen subjektiargumenttien realisoituminen joko pronominina tai nollasubjektina voidaan nähdä kaksiluokkaisena luokittelutehtävänä.

Ohjatussa oppimisessa algoritmi pyrkii oppimaan mallin, jonka avulla on mahdollista tunnistaa aineistossa esiintyvät luokat piirteiden perusteella. Täten menetelmää on mahdollista hyödyntää itse piirteiden tarkasteluun, sillä ohjelma kertoo, kuinka hyvin valitut piirteet ennustavat luokkia. Esimerkiksi Ivaska (2015) tarkastelee väitöskirjassaan suomenoppijoiden kirjoittamien tekstien morfosyntaktisia **avainrakenteita** ja Kyröläinen (2013) taas venäjän refleksiiviverbirakenteiden tyypillisiä semanttisia ja syntaktisia ominaisuuksia. Samoin avainsana-analyysi (ks. luku 3.1.) voidaan muuttaa luokitteluongelmaksi, jossa tavoitteena on oppia tunnistamaan tutkimusaineisto ja verrokkiaineisto. Luokittelun tuotoksena syntyy myös lista aineistojen tunnuspiirteistä, joita voidaan pitää luokkien avainsanoina.

Koneoppimisen hyödyntäminen avainsana-analyysissa antaa mahdollisuuksia välttää useita perinteisen, frekvenssiin perustuvan avainsana-analyysin ongelmista. Ensinnäkin koneoppimisessa tavoitteena on tunnistaa jokainen aineiston yksikkö, esimerkiksi jokainen rekisteriä edustava teksti tai tutkimuskohteena olevan kielen rakenteen käyttökerta. Näin ollen koko tutkittavaa aineistoa ei käsitellä yhtenä kokonaisuutena, eivätkä yksittäisten tekstien tai rakenteiden kohdalla tapahtuvat erot samalla tavalla vääristä tuloksia. Esimerkiksi yhdessä aineiston tekstissä erityisen taajaan esiintyvä sana ei todennäköisesti tällä menetelmällä päädy avainsanaksi. Toinen koneoppimisen selkeä etu on, että luokittelijan toimintavarmuutta pystytään arvioimaan erittäin tarkasti saannin ja tarkkuuden avulla (Tieteen termipankki s.v. saanti, tarkkuus; ks. myös kuvio 3 alla). Tämän ansiosta myös avainsanojen luotettavuutta voidaan tarkastella varsin yksityiskohtaisesti. Näin

mitattuna toimintavarmuuteen ei myöskään vaikuta aineiston koko, joten menetelmä soveltuu hyvin myös suurten aineistojen tarkasteluun. Koneoppimisen soveltaminen avaimuuden tarkasteluun on hyödyllistä myös siksi, että se on menetelmällisesti erittäin joustava ja sen avulla voidaan itse avainsanojen lisäksi tarkastella myös muita piirteitä, kuten lemmoja tai syntaktisia rakenteita. Tätä osa-aluetta havainnollistamme tarkemmin luvussa 4.3.

Ohjattua koneoppimista ja luokittelua suorittavia menetelmiä on luonnollisesti useita. Kielentutkimuksessa kenties eniten on sovellettu satunnaismetsä-nimistä menetelmää (engl. *random forest*; Tagliamonte & Baayen 2012; Kyröläinen 2013; Ivaska 2015). Toinen usein käytetty algoritmi on luvussa 4.3. esittelemämme tukivektorikone (engl. *support vector machine*). Näille algoritmeille on olemassa valmiita kirjastoja sekä R-ympäristössä että ohjelmointikieli Pythonissa.

### 3.3.3. OHJAAMATON KONEOPPIMINEN

Ohjatun koneoppimisen lisäksi toinen yleinen koneoppimisen tyyppi on ohjaamaton koneoppiminen. Ohjaamattoman koneoppimisen eri lähestymistavoista käsittelemme tässä klusterointia ja topiikkimallintamista (engl. *topic modeling*). Toisin kuin ohjatussa koneoppimisessa klusteroinnissa ja topiikkimallintamisessa ei ole valmiita luokkia, joihin aineistoa jaoteltaisiin (ks. Divjak & Fieller 2014; Moisl 2015). Ohjaamatonta koneoppimista hyödynnetään tyypillisesti silloin, kun tavoitteena on löytää aineistosta ennen tuntemattomia suhteita piirteiden avulla. Ohjaamaton koneoppiminen ja mallin arviointi on kuitenkin tunnetusti haasteellista. Siinä missä ohjatussa koneoppimisessa mallin toimivuutta voidaan arvioida sen perusteella, kuinka hyvin se tunnistaa etukäteen määritetyt luokat, ohjaamattomassa koneoppimisessa etukäteen tunnettuja luokkia ei ole.

Ohjaamattoman koneoppimisen tuloksia arvioidaan tyypillisesti kahdella kriteerillä. Yhtenä kriteerinä on opittujen luokkien evaluoiminen jollakin objektiivisella mittarilla, esimerkiksi vakauden perusteella. Tällä tarkoitetaan sitä, että mallin aineistosta oppimat luokat eivät radikaalisti muutu, vaikka piirteiden frekvenssit aineistossa vaihtelisivat hiukan. Toisena kriteerinä on opittujen luokkien selitysvoimaisuus. Tämä

kriteeri on subjektiivinen: jos mallin oppimien luokkien avulla on mahdollista selittää tutkittavan ilmiön ominaisuuksia, on mahdollista, että luokat edustavat tutkittavaa ilmiötä (ks. esim. Laippala ja muut 2018).

Yksi tavallisimmista ohjaamattomista koneoppimismenetelmistä on **klusterointi**. Klusteroinnissa tavoitteena on jakaa aineistossa esiintyvät objektit piirteiden perusteella ryhmiin, joita kutsutaan *klustereiksi*. Useimmat klusterointialgoritmit perustuvat siihen, että ne pyrkivät jakamaan aineistossa esiintyvät objektit ryhmiin siten, että tiettyyn ryhmään kuuluvat objektit ovat mahdollisimman samankaltaisia keskenään ja toisaalta samaan ryhmään kuuluvat objektit ovat mahdollisimman erilaisia suhteessa muihin aineistossa esiintyviin ryhmiin. Klusterianalyysissa tutkittavat objektit voivat käytännössä edustaa mitä tahansa kielen elementtejä aina morfologisista piirteistä ja lauseista teksteihin. Esimerkiksi Divjak ja Gries (2006) tarkastelevat venäjän yrittämistä ilmaisevien verbien merkitysten samankaltaisuuksia ja eroavaisuuksia klusteroimalla. Heidän työssään verbien lausetason käyttökontekstit on piirteistetty semanttisin, morfologisin ja syntaktisin perustein. Berez ja Gries (2009) taas soveltavat klusterointia ja samantyyppistä piirteistystä tarkastellakseen englannin *get* 'saada' verbin eri merkityksiä. Laippala ja muut (2018) puolestaan hyödyntävät klusterointia suomen kielen konnektorien, kuten *toisaalta* ja *niinpä*, tarkastelemiseen.

Toinen tavallinen ohjaamaton koneoppimismenetelmä on **topiikkimallintaminen**. Siinä tavoitteena on mallintaa ennestään tuntemattomia aihepiirejä eli topiikkeja tekstiaineistosta (ks. Blei 2012). Toisin sanoen menetelmällä voidaan tarkastella, mitä aihepiirejä aineiston dokumentit käsittelevät. Topiikkimallintamista on toistaiseksi sovellettu enemmän muissa ihmistieteissä kuin kielentutkimuksessa. Esimerkiksi Krestel ja Mehta (2010) ovat käyttäneet sitä relevanttien uutisten tunnistamiseen viestinnän tutkimuksessa ja Quinn ja muut (2010) Yhdysvaltojen senaatissa pidettyjen puheiden aihepiirien tarkasteluun politiikan tutkimuksessa. Margaret Roberts ja hänen tutkimusryhmänsä ovat soveltaneet topiikkimalleja sosiaalitieteissä (ks. Roberts, Stewart & Tingley 2016).

Yhtenä keskeisenä erona nykyaikaisten topiikki- ja klusterimallien välillä on se, että tyypillisesti ns. kovaklusteroinnissa yksi objekti, kuten tekstidokumentti, voi kuulua samanaikaisesti vain yhteen klusteriin.

Topiikkimallissa taas yksittäinen teksti voi koostua useista eri topiikeista. Jos tavoitteena on tarkastella tekstien rakentumista aihepiirien perusteella, kovaklusterointi voi olla ongelmallista, koska yksittäiset tekstit voivat sisältää useita eri topiikkeja. Esimerkiksi sanomalehti voi koostua eri osioista, kuten urheilu, talous ja ulkomaat. Tämäntyyppiselle aineistolle topiikkimalli onkin luonteva valinta.

Käytännössä topiikkimallin peruslähtökohtana on, että topiikki muodostuu kokoelmasta sanoja, jotka esiintyvät yhdessä samassa tekstissä mutta eivät välttämättä esiinny lineaarisesti yhdessä tekstitasolla. Teknisemmin ajateltuna jokainen teksti on jakauma eri topiikkeja ja jokainen topiikki on jakauma sanoja, joita koko aineistoon kuuluu. Tiivistäen topiikkimallintamisen avulla pystytään 1) oppimaan koko tekstiaineiston topiikkirakenne, 2) arvioimaan yksittäisten tekstidokumenttien topiikkirakenne ja 3) kuvaamaan aineiston topiikit avainsanojen perusteella. Menetelmä antaa mahdollisuuden tarkastella tekstiaineistossa esiintyviä aihepiirejä toisaalta monipuolisemmin ja toisaalta yksityiskohtaisemmin kuin esimerkiksi perinteinen avainsana-analyysi. Viimeksi mainittuun verrattuna erona on myös, että topiikkimalli pyrkii oppimaan aineistossa esiintyvät topiikit sen perusteella, miten sanat ovat jakautuneet aineiston sisällä vertaamatta aineistoa toiseen aineistoon.

Määrällisen korpuslingvistiikan näkökulmasta tutkimus ei välttämättä ole rajoittunut tarkastelemaan ainoastaan tietyn tekstin aihepiirejä, vaan erityisesti kielitieteellisesti painottuneelle tutkimusaiheelle saattaa olla myös oleellista tarkastella, miten aihepiireistä on kirjoitettu. Tutkimuksessa voidaan esimerkiksi tarkastella, miten tekstin tietty ominaisuus, kuten syntaktinen kompleksisuus, tai jokin muu metatieto korreloi tiettyjen topiikkien kanssa. Perinteinen topiikkimalli ei kuitenkaan suoraan sovellu tähän. Tätä varten Roberts, Stewart, Tingley ja Airoldi (2013) ovatkin kehittäneet rakennetopiikkimallin. Se edustaa topiikkimallia, joka mahdollistaa tekstien rakenteellisten piirteiden tai muun metatiedon ottamisen huomioon topiikeissa. Algoritmi on saatavilla R-ympäristössä. Luvussa 4.4. tarkastelemme tekstiaineistoa juuri rakennetopiikkimallin avulla.

# 4. Tutkimuksen eteneminen

Tässä luvussa havainnollistetaan esimerkkitutkimuksella yllä esitettyjen menetelmien soveltamista omaan aineistoon Unix-käyttöympäristössä. Kaikki esimerkissä käytetyt ohjelmat ja aineistot löytyvät GitHubista osoitteesta https://github.com/TurkuNLP/Corpus-linguistics#corpus-linguistics.

Kuten jo mainittiin, monet esitellyistä menetelmistä ovat valmiina korpustyökaluissa. Näiden helppokäyttöisyyden mukana tulee kuitenkin myös haasteita. Valmiit ohjelmat tarjoavat varsin yksipuoliset mahdollisuudet aineiston analyysiin, sillä niiden sisältämiä toimintoja ei voi muokata eikä niihin voi lisätä toimintoja, jotka saattaisivat omassa tutkimuksessa olla tärkeitä. Lisäksi aineiston esikäsittely pitää aina tehdä ennen varsinaista analyysia riippumatta siitä, käyttääkö valmiita työkaluja vai itse tehtyjä ohjelmia. Lopuksi on korostettava, että valmiiden työkalujen käytössä on riski, ettei käyttäjä itse asiassa ymmärrä, mitä on tekemässä. Näiden syiden takia rohkaisemme lukijoita Unix-käyttöympäristön opetteluun.

Esimerkkitutkimuksen tavoitteena on analysoida köyhyyskeskusteluja Suomi24-palstalla. Miten köyhyydestä puhutaan? Mitä aiheita siihen liittyy ja mitä niistä sanotaan? Suomi24-keskustelupalstan tekstit on luovutettu tutkimuskäyttöön, ja ne ovat saatavilla Kielipankista. Tutkimusta varten valtavasta Suomi24-kokoelmasta on eristetty kaikki vuonna 2014 julkaistut kommentit, joissa mainitaan sana *köyhä* tai sen lähisynonyymi (esim. *varaton*, *rutiköyhä*) jossain taivutusmuodossa. Kommentteja on yhteensä 33 716 ja saneita (välimerkit mukaan luettuina) 3 375 171. Aineisto löytyy GitHubista nimellä koyha-kommentit-2014.txt.gz. Kyseessä on pakattu tekstitiedosto UTF8-muodossa, mikä takaa tiedoston koneluettavuuden ja soveltuvuuden eri käyttöalustoille. Muototietoja sisältäviä tiedostomuotoja, kuten docx-dokumentteja, ei kannata missään nimessä käyttää.

## 4.1. Tutkimusaineiston esikäsittely

Esikäsittely on tärkeä osa tutkimusta, ja sen merkitys korostuu, kun tutkimusaineistona on verkkoaineistojen pohjalta luotuja tekstiaineistoja. Esimerkiksi välimerkkien eristäminen sanoista mahdollistaa sanojen *köyhä* ja *köyhä?* tulkitsemisen samoiksi. Syntaksijäsennysohjelman avulla voidaan juoksevien sanojen sijaan käyttää ohjelman tunnistamia lemmoja ja muuttaa isot kirjaimet pieniksi.

Tässä tutkimuksessa aineisto on esikäsitelty vapaasti saatavilla olevalla Finnish Dep Parser -syntaksijäsennysohjelmalla. GitHubista löytyvässä aineistossa on saatavilla jokaisesta keskustelupalstakommentista sane, sen lemmamuoto sekä sanaluokka. Jokainen aineiston sane ja sen analyysit ovat omalla rivillään. Lisäksi jokaista kommenttia edeltävät sitä kuvaavat metatiedot: kommentin id-numero, julkaisuaika, keskustelun aihealue, otsikko ja kirjoittajan käyttäjätunnus. Kuviossa 1 on esitetty annotoitu esimerkki aineistosta.

```
# id: 6065527
# date: 2014-04-15T21:24:00Z
# sect: Paikkakunnat
# subsect: Pohjois-Pohjanmaa
# user: Maassa maan tavalla
Putin   Putin   PROPN
on      olla    VERB
raaka   raaka   NOUN
diktaattori     diktaattori     NOUN
,       ,       PUNCT
hän     hän     PRON
yrittänee       yrittää         VERB
vallata         vallata         VERB
Suomenkin       Suomi   PROPN
.       .       PUNCT
```

Kuvio 1. Annotoitu esimerkki aineistosta.

## 4.2. Avainsanat ja log-uskottavuus

Tässä luvussa lasketaan aineiston avainsanat perinteisellä menetelmällä, jossa kaikkia kommentteja käsitellään yhtenä kokonaisuutena ja avainsanat perustuvat kommenteissa yli- tai aliedustettuina oleviin sanoihin. Tässä tutkimuksessa tilastollinen todennäköisyys lasketaan log-uskottavuudella.

Verrokkiaineiston valinta on tärkeä tekijä avainsanojen määrittämisessä, sillä se, millaiseen aineistoon tutkittavan aineiston sanastoa verrataan, vaikuttaa keskeisesti tuloksiin eli saatuihin avainsanoihin. Jos aineistomme Suomi24-keskusteluja verrataan esimerkiksi uutisartikkeleihin, tulokset kertovat ennen kaikkea keskeisistä eroista uutisten ja internetkeskustelujen välillä. Nyt tavoitteena on kuitenkin tarkastella nimenomaan köyhyyttä käsittelevien keskustelujen erityispiirteitä ja niitä käsitteitä, joita keskustelijat liittävät köyhyyteen. Tämä tavoite saavutetaan parhaiten vertaamalla näitä keskusteluja Suomi24-keskusteluihin yleensä. Tätä varten koko Suomi24-aineistosta on tehty 30 miljoonan saneen yleisotos, jossa kommenttien sanastoa ei ole määritelty. Tiedoston nimi GitHubissa on verrokki.txt.gz. Yleensä ajatellaan, että verrokkiaineiston pitäisi olla mahdollisimman suuri ja ehdottomasti selvästi suurempi kuin tutkitun aineiston. Koska Suomi24-aineisto käsittää miljardeja sanoja, ei kuitenkaan ole järkevää käyttää sitä kokonaisuudessaan. Tämän vuoksi kymmenkertaista kokoa verrattuna tutkittuun aineistoon voidaan pitää riittävänä.

Avainsanojen laskemista varten GitHubista löytyy ohjelma nimeltä log_ll.py. Kun tarvittavat tiedostot eli koyha-kommentit-2014.txt.gz ja verrokki.txt.gz on tallennettu samaan kansioon kuin ohjelma, se laskee automaattisesti avainsanat molemmille tiedostoille. Juoksevien saneiden sijaan ohjelma käyttää kuitenkin tiedostojen toisessa sarakkeessa olevia lemmamuotoja. Näin aineistosta saadaan tihempää. Toisin sanoen erilaisia muotoja on vähemmän, kun esimerkiksi saneet *oli*, *lienee*, *olivat* ja *olen* on muutettu lemmaksi *olla*. Lemmojen käyttö on hyvä esimerkki itse tehtyjen ohjelmien joustavuudesta, sillä esimerkiksi AntConcin kanssa lemmojen käyttö saneiden sijaan ei ole yhtä helppoa.

Kuvio 2 alla näyttää otoksen ohjelman tuotoksesta. On tärkeä huomata, että ohjelma poistaa aineistoista automaattisesti köyhyyteen liittyvät hakusanat, joita kommenttien eristämiseen koko aineistosta käytettiin. Tuotoksessa avainlemmat näkyvät tärkeysjärjestyksessä, jokainen lemma omalla rivillään. Ohjelma tuottaa sekä positiiviset eli tutkitussa aineistossa yliedustetut että negatiiviset eli aineistossa aliedustetut avainsanat. Jokaiselle lemmalle ohjelma tuottaa 1) log-uskottavuusarvon, 2) frekvenssin köyhät-aineistossa ja koko aineistossa yhteensä, 3) tiedon siitä, onko lemma positiivinen vai negatiivinen, sekä 4) itse lemman. Kuviossa 2 tärkeimmät positiiviset lemmat ovat *rikas*, *raha*, *[]*, *työtön*, *eläke*, *maksaa*, *Kannbis*, *köyhyys* ja *,*.

```
Suomi24 koyhat
ll=26932.342   FREQ=36/68342   Neg   kita
ll=17194.163   FREQ=11003/160024   Neg   !
ll=12516.490   FREQ=191/35668   Neg   auki
ll=10665.729   FREQ=765/39653   Neg   kiinni
ll=9501.870   FREQ=5032/7059   Pos   rikas
ll=8364.628   FREQ=10512/23977   Pos   raha
ll=5271.378   FREQ=34/14064   Neg   tuputtaa
ll=4335.555   FREQ=3328/5868   Pos   työtön
ll=4187.831   FREQ=2997/5071   Pos   eläke
ll=4049.368   FREQ=8827/25143   Pos   maksaa
ll=3882.584   FREQ=1139/1139   Pos   []
ll=3152.283   FREQ=187413/914719   Pos   ,
ll=2971.648   FREQ=7347/21832   Pos   työ
ll=2883.816   FREQ=846/846   Pos   Kannbis
ll=2775.724   FREQ=2214/3991   Pos   kokoomus
ll=2653.797   FREQ=1403/1966   Pos   köyhyys
ll=2599.041   FREQ=4780/12799   Pos   euro
ll=2581.906   FREQ=2671/5542   Pos   vara
```

Kuvio 2. Köyhät-aineiston avainlemmoja log_ll.py-ohjelmalla laskettuna.

Positiivisista avainlemmoista monet kertovat aineiston aihepiiristä: on loogista, että köyhyyttä käsittelevissä keskusteluissa aihepiireinä ovat myös raha, rikkaat ja eläkeläiset. Sanan *Kannbis* läsnäolo listalla on hyvä esimerkki ongelmista, joita syntyy, kun avainsanat lasketaan aineistosta kokonaisuutena. Sanetta on käytetty ilmeisesti koneellisesti tuotetuissa kommenteissa. Itse kommentit ovat täsmälleen identtisiä, mutta ne on julkaistu eri aikoina. Tämän takia ne ovat päätyneet myös aineistoon erillisinä yksiköinä. Kommentteja on yhteensä 18, ja niissä sana *Kannbis*

mainitaan yhteensä 846 kertaa. Tämäntyyppisten ”tulosten” välttämiseksi on mahdollista esikäsitellä aineistoa tarkemmin.

Toinen merkittävä epäkohta kuvion 2 avainlemmoissa on välimerkkien läsnäolo. Kun tutkimusaiheena on merkityksen tarkastelu isossa tekstiaineistossa, välimerkeillä on minimaalinen kontribuutio merkityksen muodostumiseen. Ne olisi siten hyvä poistaa jo esikäsittelyvaiheessa. GitHubissa on ohjelma log_likelihood_content_words.py, joka laskee aineistosta avainlemmat käyttäen hyväkseen ainoastaan sanoja, joiden sanaluokka on substantiivi (vain yleisnimet), erisnimi, verbi tai adverbi. Voit itse koettaa ohjelmaa ja katsoa, miten tulokset muuttuvat!

## 4.3. Avainsanat ja ohjattu koneoppiminen

Luvussa lasketaan avainsanat köyhät-aineistolle ohjatun koneoppimisen avulla. Ohjelmana käytetään lineaarista tukivektorikonetta. Koneoppimisluokittelijoita on monia erilaisia. Tukivektorikone on erittäin paljon käytetty, ja se soveltuu moniin erilaisiin luokittelutehtäviin kielitieteestä biologiaan. Lisäksi se on valmiina monissa eri kirjastoissa ja paketeissa eri alustoille. Esimerkiksi Python-ohjelmointikielen Scikit-learn-kirjasto[3] sisältää tukivektorikoneen ja monia muita koneoppimismenetelmiä.

Tässä tehtävässä verrokkiaineistona käytetään Suomi24-alueella julkaistuja kommentteja, joissa ei mainita lemmaa *köyhä* tai sen lähisynonyymejä. Kaikista vuonna 2014 julkaistuista kommenteista otettiin tätä varten samankokoinen otos kuin köyhät-aineistosta. Luokittelijan toiminnan arviointi on suoraviivaisempaa, kun luokiteltavat aineistot ovat samankokoisia. Tällöin tiedetään, että 50 %:n tarkkuuden saavuttanut luokittelija ei oppinut aineistossa esiintyviä luokkia, sillä tämän arvon voi saavuttaa luokittelemalla kaikki kommentit toiseen luokkaan. Lisäksi luokittelijan harjoitusaineiston kokoamisessa on tärkeää, että luokat pysyvät erillään. Jos harjoitusaineistoon päätyy esimerkiksi kommentteja, joihin on merkitty väärä luokka, luokittelija ei välttämättä pysty oppimaan aineiston rakennetta.

GitHubissa tukivektorikoneen sisältämä ohjelma löytyy nimellä svm.py. Ohjelma ottaa syötteekseen rivitettyä aineistoa, jossa jokainen luokiteltava objekti eli tässä tapauksessa Suomi24-aineistosta peräisin oleva

kommentti on omalla rivillään. Jokainen rivi alkaa kommenttiluokan kategorialla, jonka tulee olla numeerisessa muodossa. Tätä varten GitHubissa on myös tiedosto preprocess.py, joka muuttaa köyhät-aineiston ja verrokkiaineiston vaadittuun muotoon ja myös numeroi rivit. Köyhät-aineistosta peräisin olevat kommenttirivit alkavat numerolla 0 ja verrokkiaineiston numerolla 1. Lisäksi preprocess.py poistaa kommenteista sellaisiin sanaluokkiin kuuluvia sanoja, joiden ajatellaan tuovan vain vähän lisäarvoa tai jopa häiritsevän analyysia. Konjunktiot ovat hyvä esimerkki tällaisista sanoista. Tässä esimerkissä on konjunktioiden lisäksi poistettu apuverbit, pronominit, välimerkit sekä rinnastus- ja alistuskonjunktiot. Tarkemmat ohjeet ohjelmien ajamiseen löytyvät GitHubista.

Kuvio 3 alla esittää kuvakaappauksen luokittelijan tuotoksesta. Ensin molemmille kommenttiluokille, köyhät-aineistolle (0) ja verrokkiaineistolle (1), on laskettu **saanti** (engl. *recall*), **tarkkuus** (engl. *precision*) ja näiden yhdistelmä, ns. f1-arvo. Saanti kuvaa sitä, kuinka suuren osan kaikista luokan kommenteista luokittelija löysi. Tarkkuus taas kuvaa sitä, kuinka suuri osa löydetyistä kommenteista oli oikein. Support-sarake oikealla näyttää kommenttien määrän harjoitusaineistossa kummallekin luokalle erikseen ja koko aineistossa yhteensä.

```
             precision    recall  f1-score   support

          0       0.88      0.87      0.88      6738
          1       0.87      0.89      0.88      6725

avg / total       0.88      0.88      0.88     13463


 CLASS 0: Positiiviset avainlemmat:

  rikas -1.06491209855
  köyhyys -0.868217517019
  vara -0.531522030804
  elättää -0.51808762324
  raha -0.436919562097
  omaisuus -0.42775190325
  äänestää -0.404974030735
  työtön -0.338692615464
  ilmainen -0.33713834136
  veron#maksaja -0.300362858592

 CLASS 1: Positiiviset avainlemmat:

  kokemus 0.124108203712
  pyytää 0.126407954952
  tapahtua 0.133383090873
  hoito 0.13911178754
  luoda 0.139132227712
  tulos 0.141697125662
  viesti 0.143187313088
  väittää 0.143485714644
  tila 0.173655177397
  yhteys 0.217014277449
```

Kuvio 3. Luokittelijan suorituskyky ja avainlemmat. Luokka 0 on köyhät-aineisto, luokka 1 verrokkiaineisto.

Luokittelijan f1-arvon keskiarvo on 0.88 eli 88 %. Toisin sanoen keskimäärin 88 % kommenteista on tunnistettu oikeaan luokkaan. Tämä on ihan kelvollinen tulos. Se kertoo myös, että luokat ja niiden lemmat ovat oikeasti erilaisia. Jos luokkien lemmojen välillä ei olisi merkittäviä eroavaisuuksia, luokittelija ei pystyisi näin hyvin tunnistamaan niitä. Itse avainsanat antavat hyvän kuvan kommenttien sisällöstä. Verrattuna edellisen luvun log-uskottavuudella laskettuihin avainsanoihin suurimmat erot näyttäisivät olevan välimerkkien ja muiden hyödyttömien avainsanojen, kuten *Kannbis,* puuttumisessa.

## 4.4. Topiikkimallintaminen

Tässä luvussa tarkastelemme köyhät-aineistoa topiikkimallin avulla hyödyntäen rakennetopiikkimallia, joka on saatavilla stm-paketissa R-ohjelmistolle. Tavoitteena on laskea aineiston topiikit ja arvioida, kuinka jokin metatieto, esimerkiksi tekstin rakenne, voi vaikuttaa topiikkeihin. Tässä artikkelissa keskitymme tarkastelemaan kommentin kirjoituskuukauden ja sanapituuden vaikutusta topiikkeihin. Ensimmäisessä luvussa esittelemme aineiston muokkausta R-ympäristössä ja toisessa varsinaista topiikkimallintamista.

### 4.4.1. TEKSTIAINEISTON MUOKKAUS TOPIIKKIMALLINTAMISEEN R-YMPÄRISTÖSSÄ

Tässä luvussa esittelemme yleisellä tasolla aineiston esikäsittelyn topiikkimallintamista varten. Koska topiikkimallin tavoitteena on oppia aineistossa esiintyvät aihepiirit, on tärkeää, että tekstiaineisto on esikäsitelty perusteellisesti. Tätä varten GitHubissa on köyhät-aineisto, jonka tekstiaines on rajattu tiettyihin sanaluokkiin: verbi, substantiivi (vain yleisnimet), erisnimi, adjektiivi ja adverbi. Tämä valmiiksi esikäsitelty aineisto on saatavilla artikkelin Github-sivustolla (dat_stm), ja sitä voi käyttää sellaisenaan topiikkimallintamisessa.

Varsinaista topiikkimallintamista varten aineisto on muutettava mallin olettamaan muotoon, sanasäkkiin. Tekstin muokkauksessa olemme hyödyntäneet quanteda R -pakettia (versio 1.3.13), joka tarjoaa kattavan

tuen keskeisille korpuslingvistiikan operaatioille ja jonka käyttö on myös hyvin dokumentoitu. Analyysia varten tekstiaineisto on ensin muutettu korpukseksi. Näin saamme sekä varsinaisen tekstiaineiston että siihen liittyvät piirteet nivotuksi yhteen. Se käy komennolla

```
corpus2014 <- corpus(dat$text, docnames = dat$doc, docvars=
select(dat, doc, month, Nwords_Log, nick, section, subsection))
```

Luotu korpus koostuu varsinaisesta tekstiaineistosta (dat$text), dokumenttien nimistä (dat$doc) ja muuttujista (dat, doc, month, Nwords_Log, nick, section, subsection). Korpus-objektiin liittyviä komponentteja voi tarkastella komennolla

```
summary(corpus2014)
```

Yksittäisiä kommentteja voi tarkastella komennolla texts. Esimerkiksi kommenttia 2 voi tarkastella komennolla

```
texts(corpus2014)[2]
```

Tekstiaineisto muutetaan sanasäkkimuotoon, jota kutsutaan myös *dokumentti-termi-matriisiksi* tai *dokumentti-piirre-matriisiksi*. Köyhät-aineistossa kommentit vastaavat dokumentteja. Tässä muodossa tekstiaineiston dokumentit edustuvat riveinä ja termit eli sanat ovat sarakkeina. Matriisin solut vastaavat puolestaan frekvenssiä. Dokumentti-termi-matriisi luodaan komennolla dfm(). Lisäksi poistamme mahdolliset välimerkkien ja symbolien jäänteet teksteistä argumenteilla remove_punct ja remove_symbols.

```
dat_dfm <- dfm(corpus2014, remove_punct = TRUE, remove_symbols = TRUE)
```

Voimme tarkistaa, että matriisi muodostui oikein komennolla dim(dat_dfm), joka laskee matriisin ulottuvuudet. Komento palauttaa arvot 32 600 ja 143 120, jotka vastaavat rivien ja sarakkeiden määriä

matriisissa. Aineisto on siis oikeassa muodossa, koska aineistossa esiintyy kommentteja 32 600 ja sanoja 143 120. Lisäksi poistamme aineistosta sanat, joiden taajuus oli alle 5 ja jotka esiintyivät alle viidessä kommentissa.

```
dat_dfm_trim <- dfm_trim(dat_dfm, min_count = 5, min_docfreq = 5)
```

Viimeisenä muokkauksena muutamme vielä dokumentti-termi-matriisin rakennetopiikkimallin olettamaan muotoon, ja samalla tuomme dokumenttien metatiedot ja kielellisen rakenneinformaation aikaisemmin luodusta korpuksesta:

```
dat_stm <- convert(dat_dfm_trim, to = ”stm”, docvars = docvars(corpus2014))
```

Olemme nyt saattaneet aineiston sellaiseen muotoon, että voimme sovittaa siihen rakennetopiikkimallin. Tämä menetelmä esitellään seuraavassa luvussa.

### 4.4.2. TEKSTIAINEISTON TOPIIKKIMALLINTAMINEN

Rakennetopiikkimalli sovitetaan aineistoon R:ssä komennolla stm(). On kuitenkin syytä mainita, että kyseessä on suurehko aineisto ja mallin sovittaminen vaatii useita tunteja tehokkaallakin tietokoneella. Käsitelty aineisto ladataan R:n istuntoon seuraavalla komennolla:

```
load(”dat_stm.Rda”)
```

Tätä artikkelia varten sovitimme mallin seuraavalla komennolla:

```
stm_evex <- stm(dat_stm$documents, dat_stm$vocab, K = 0, init.type = ”Spectral”, prevalence = ~ s(Nwords_Log) + s(month, df = 4), data = dat_stm$meta, seed = 12345)
```

Rakennetopiikkimallintamista varten komento kattaa kaikki keskeiset osa-alueet. Ensin on määritelty aineistossa esiintyvät kommentit (dat_stm$documents) ja sanasto (dat_stm$vocab), joiden perusteella malli laskee topiikit. Koska kyseessä on rakennetopiikkimalli, kommennossa on mukana prevalence-argumentti. Tämän avulla voimme ottaa topiikkien estimoinnissa huomioon, että topiikkien esiintymismäärään saattaa vaikuttaa jokin muuttuja. Tässä analyysissa otamme huomioon kommentin sanapituuden ilmaistuna log-yksikköinä (Nwords_Log) ja kommentin kirjoitusajankohdan ilmaistuna kuukautena (month).

Rakennetopiikkimalli ei oleta, että muuttujien, kuten sanapituuden tai kuukauden, välinen suhde olisi lineaarinen topiikkien esiintymisfrekvenssin kanssa. Ajallisessa aineistossa sanojen frekvenssi ei tyypillisesti ole vakio vaan niiden käyttö vaihtelee. Jotta voimme mallintaa frekvenssin ja ajan suhdetta mahdollisimman tarkasti, on tärkeää ottaa huomioon mahdollinen ajan ja frekvenssin välisen suhteen epälineaarisuus. Kattava johdatus näihin malleihin R-ympäristössä on esitetty Simon Woodin teoksessa *Generalized Additive Models: An Introduction with R* (2017).

Funktio s() sisältää argumentin df, joka kontrolloi muuttujan epälineaarisuutta. On myös mahdollista sovittaa lineaarinen malli, jolloin funktio s() jätetään pois kommennosta. Oletusarvoisesti df saa arvon 10, mutta kuukausi-muuttujan kohdalla arvo on laskettu, koska kuukausi sisältää vain 12 uniikkia arvoa. Lisäksi prevalence-argumentti vaatii metatietojen lähteeksi objektin, joka spesifioidaan argumentilla data = dat_stm$meta.

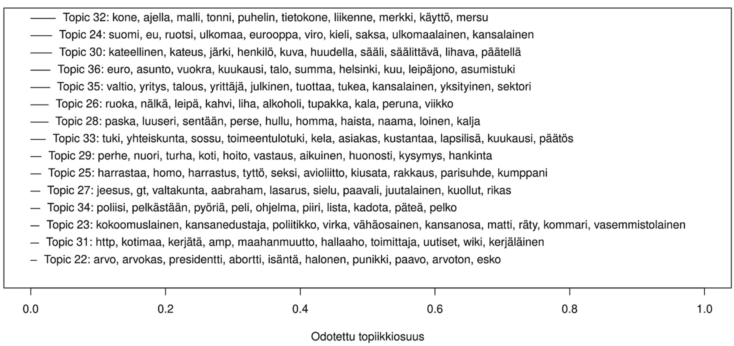
On syytä muistaa, että kyseinen s-funktio edustaa aineistovetoista menetelmää, jossa kuukauden tai sanapituuden epälineaarisuus estimoidaan aineistosta (ks. Wood 2017). Topiikkimallien kohdalla ei ole yksiselitteistä menetelmää, jonka avulla voisi valita optimaalisen arvon df-argumentille. Tässä tapauksessa valitsemme arvon 4, koska se heijastaa ainakin osittain kommenttien epälineaarista taajuutta aineistossa: Tammikuussa oli yhteensä 1 382 kommenttia, mutta kommenttien taajus notkahti helmikuussa, jolloin niitä oli 1 282. Tämän jälkeen kommenttien taajuus lähti kasvuun ja saavutti yhden kasvupisteen huhtikuussa (4 202 kommenttia), jonka jälkeen kommenttien määrä lähti selvään laskuun saavuttaen matalimman pisteen lomakuukautena heinäkuussa (464). Lopuksi kommenttien määrä lähti taas nousemaan

saavuttaen huipun joulukuussa (4 722). Mallintamisemme tavoitteena on siten sovittaa aineistoon tarpeeksi joustava malli, jotta se heijastaisi aineistoa mahdollisimman tarkasti. On kuitenkin syytä muistaa, että topiikkimallintamisen yhteydessä tavoitteenamme on tarkastella, miten mallin estimoimat topiikit korreloivat tietyn muuttujan kanssa.

Topiikkimallintamisen yhtenä haasteena on, että mallit olettavat, että topiikkien määrä (*k*) on etukäteen määritelty. Tässä artikkelissa käytetty rakennetopiikkimalli pyrkii arvioimaan aineiston perusteella topiikkien määrän. Tämä menetelmä toimii parhaiten isojen aineistojen kanssa ja sitä hyödynnetään valitsemalla k-argumentille arvo nolla ja aloitusmenetelmäksi Spectral (init.type = ”Spectral”). Lopuksi on syytä muistaa, että topiikkimallit ovat stokastisia eli algoritmin tulokset saattavat erota ajokertojen välillä. Tämän vuoksi komennossa on argumentti seed, joka spesifioi satunnaisgeneraattorin aloituksen.

Rakennetopiikkimalli estimoi, että aineisto koostuu 46 topiikista. Keskeinen kysymys on tietysti, mitä tuloksena saadut topiikit tarkoittavat. R-paketti stm sisältää useita valmiita funktioita, joiden avulla on mahdollista tarkastella rakennetopiikkimallin tuloksia. Topiikkien arviointia varten voidaan laskea ja visualisoida niiden avainsanat. Tämä tapahtuu seuraavalla komennolla (ks. kuvio 4):

```
plot(stm_evex, type = ”summary”, n = 25, xlim = c(0, 1), labeltype = ”prob”, topics = c(22:36))
```




Kuvio 4. Topiikkien 22–36 estimoidut avainsanat.

Avainsanojen perusteella voi ensinnäkin todeta, että malli näyttää varsin onnistuneelta. Monien topiikkien avainsanoista voi hyvin päätellä niiden edustaman aihepiirin. Esimerkiksi topiikin 26 avainsanat *ruoka*, *nälkä*, *leipä*, *kahvi*, *liha*, *alkoholi* ja *tupakka* kertovat selvästi, mistä tässä topiikissa puhutaan. Topiikin 28 alatyyliset ja halventavat avainsanat taas kertovat köyhyyteen liittyvistä asenteista ja eriarvoisuudesta.

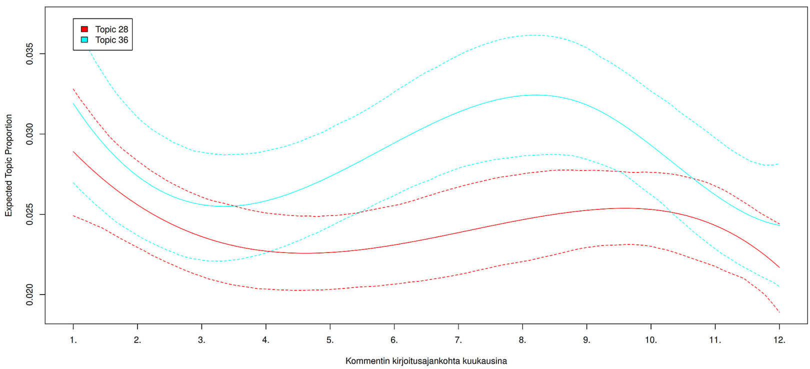
Avainsanojen ohella rakennetopiikkimallista voi tutkia, miten muuttujat vaikuttavat topiikkien esiintymisfrekvenssiin. Ensin estimoimme muuttujan ja kaikkien topiikkien välisen suhteen komennolla

```
prep <- estimateEffect(1:46 ~ s(Nwords_Log) + s(month, df = 4),
stm_evex, meta = dat_stm$meta)
```

Funktiossa estimateEffect() spesifioidaan ensin kaikki topiikit (1:46) ja sitten muuttujat. Tämän jälkeen on mahdollista verrata tiettyjä topiikkeja keskenään. Tässä tapauksessa käytämme muuttujina kommentin kirjoitusajankohtaa ja vertaamme keskenään topiikkeja 28 ja 36 komennolla:

```
plot(prep, ”month”, method = ”continuous”, topics = c(28, 36))
```

Topiikin esiintymisfrekvenssin ja kirjoitusajankohdan korrelaatio on visualisoitu kuviossa 5.




Kuvio 5. Kommentin kirjoitusajakohdan vaikutus topiikkien 28 ja 36 esiintymisfrekvenssiin.

Topiikin 28 esiintymisfrekvenssi ei paljoakaan vaihtele kuukausittain, ja kommenttien määrä näyttää pysyvän suhteellisen tasaisena. Topiikkissa 36 puolestaan näkyy esiintymien selvä sesonkiluonteisuus, sillä kommenttien määrä on suurempi heinä- ja elokuussa.

Olemme esitelleet lyhyesti tässä artikkelissa topiikkimallin ja sen yhden sovelluksen R-ympäristössä. Rakennetopiikkimalli on hyvin dokumentoitu, ja se sisältää kattavan määrän funktioita, joilla voi tarkastella tekstiaineistoja. Rakennetopiikkimalli avaa monipuolisen näkymän jäsentämättömään tekstiaineistoon, ja tässä osuudessa olemme käyneet läpi sen keskeisimmät ominaisuudet kielellisen analyysin kannalta.

# 5. Yhteenveto

Tässä artikkelissa olemme esitelleet määrällistä korpuslingvistiikkaa ja sen menetelmiä. Määrällisessä korpuslingvistiikassa tyypilliset tutkimuskysymykset liittyvät joko muotojen variaatioon tai merkitysten rakentumiseen muotojen kautta. Tutkimusparadigmana määrällinen korpuslingvistiikka yhdistää kielitiedettä ja laskennallisia menetelmiä tilastollisista menetelmistä koneoppimiseen. Menetelmällisen rikkauden ansiosta määrällinen korpuslingvistiikka tarjoaa kiehtovan näkökulman kieleen.

Korpuslingvistiikkaan yhdistyy läheisesti tietotekniikka. Tässä luvussa olemme näyttäneet myös käytännössä, miten tietotekniikkaa voidaan soveltaa korpuslingvistiikan kysymyksiin. Lisäksi tietotekniikan kehitys pitää korpuslingvistiikan jatkuvassa liikkeessä, sillä uusia menetelmiä ja mahdollisuuksia syntyy koko ajan. Tästä syystä mekin olemme esitelleet erilaisia koneoppimismenetelmiä. Vaikka niitä ei välttämättä vielä käytetä laajasti, voidaan olettaa, että tulevaisuudessa niiden suosio kasvaa. Esimerkiksi topiikkimallintamisella on laajasti sovellusmahdollisuuksia kielentutkimuksessa.

**Työkaluja**

Korpustyökaluja:

- AntConc: http://www.laurenceanthony.net/software/antconc/
- CQPweb: http://www.research.lancs.ac.uk/portal/en/upmprojects/cqpweb(855e22e7-97fb-4863-9101-9b245729ccbd).html
- VARD 2 -työkalu historiallisten tekstien kirjoitusasun standardisointiin: http://ucrel.lancs.ac.uk/vard/
- WordSmith Tools: http://www.lexically.net/wordsmith/
- R:n NLP paketteja: https://cran.r-project.org/web/views/NaturalLanguageProcessing.html

Laskennallisia työkaluja:

- SPSS: https://www.ibm.com/analytics/spss-statistics-software/
- VARBRUL: https://en.wikipedia.org/wiki/Variable_rules_analysis

Ohjelmointikieliä:

- Python: https://www.python.org/
- R-project: https://www.r-project.org
- ggplot2: https://ggplot2.tidyverse.org (monipuolinen grafiikkapaketti R:ään)
- tidyverse: https://www.tidyverse.org/ (R-pakettikokoelma aineiston muokkaamiseen, mallintamiseen ja visualisointiin)
- Scikit-learn -kirjasto: http://scikit-learn.org/stable/

Työkaluja puhutun aineiston käsittelyyn:

- ELAN: https://tla.mpi.nl/tools/tla-tools/elan/
- Praat: http://www.fon.hum.uva.nl/praat/

Työkaluja typologiseen ja historiallis-vertailevaan tutkimukseen:

- Max Planck Institute for Evolutionary Anthropology, Department of Linguistics: https://www.eva.mpg.de/lingua/tools-at-lingboard/tools.php

- DatSemShifts, Database of Semantic Shifts: http://datsemshift.ru
- Database of cross-linguistic colexifications: https://clics.clld.org
- Leipzigin glossausohjeet: https://www.eva.mpg.de/lingua/resources/glossing-rules.php

**Aiheesta lisää:**

Biber, Douglas & Reppen, Randi. 2015. *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press.

Bird, Steven, Klein, Ewan & Loper, Edward. 2009. *Natural Language Processing with Python: {A}analyzing Text with the Natural Language Toolkit*. Sebastopol: O'Reilly Media, Inc.

Jockers, Matthew L. 2014. *Text Analysis with R for Students of Literature*. New York (NY): Springer.

Manning, Christopher, Raghavan, Prabhakar & Schütze, Hinrich. 2008. *Information Retrieval*. Cambridge: Cambridge University Press.

## VIITTEET

1 https://www.kielipankki.fi/aineistot/oulu/
2 Tässä termi *diskurssi* määritellään näin: "a set of meanings, metaphors, representations, images, stories, statements and so on that in some way together produce a particular version of events" (Burr 1995, 48).
3 http://scikit-learn.org/stable/.

## KIRJALLISUUS

Aller Media Oy. 2014. The Suomi 24 Sentences Corpus (2016H2). [tekstikorpus]. Kielipankki. Saatavissa: http://urn.fi/urn:nbn:fi:lb-2017021505.

Anthony, Laurence. 2014. AntConc (Windows, Macintosh OS X, and Linux) Build 3.4.3. Saatavissa: https://www.laurenceanthony.net/software/antconc/releases/AntConc343/help.pdf.

Arppe, Antti. 2008. *Univariate, Bivariate and Multivariate Methods in Corpus-based Lexicography: a Study of Synonymy*. Väitöskirja. Helsinki: University of Helsinki.
Baker, Paul, Gabrielatos, Costas, KhosraviNik, Majid, Krzyżanowski, Michał, McEnery, Tony & Wodak, Ruth. 2004. A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse and Society* 19:3, 273–306.
Berez, Andrea L. & Gries, Stefan Th. 2009. In defense of corpus-based methods: a behavioral profile analysis of polysemous *get* in English. Julkaisussa: Moran, Steven, Tanner, Darren S. & Scanlon, Michael (toim.) *Proceedings of the 24th Northwest Linguistics Conference*. University of Washington Working Papers in Linguistics Vol. 27. Seattle (WA): University of Washington, Department of Linguistics, 157–166.
Biber, Douglas. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.
Biber, Douglas & Conrad, Susan. 2009. *Register, Genre, and Style*. Cambridge: Cambridge University Press.
Blei, David M. 2012. Probabilistic topic models. *Communications of the ACM* 55:4, 77–84.
Burr, Viven. 1995. *An Introduction to Social Constructionism*. London: Routledge.
Bybee, Joan L. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 5, 425–455.
Bybee, Joan L. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
Cuskley, Christine F., Pugliese, Martina, Castellano, Claudio, Colaiori, Francesca, Loreto, Vittorio & Tria, Francesca. 2014. Internal and external dynamics in language: Evidence from verb regularity in a historical corpus of English. *PloS ONE* 9:8, e102882.
Divjak, Dagmar S. & Gries, Stefan Th. 2006. Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2:1, 23–60.
Divjak, Dagmar & Fieller, Nick. 2014. Clustering linguistic data. Julkaisussa: Glynn, Dylan & Robinson, Justyna (toim.) *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*. Amsterdam: John Benjamins, 405–441.
Firth, John Rupert. 1957. *Papers in Linguistics 1934–1951*. London: Oxford University Press.
Fries, Charles C. 1940. *American English Grammar: The Grammatical Structure of Present-day American English with Special Reference to Social Differences or Class Dialects*. New York (NY) & London: D. Appleton-Century Co.
Harris, Zellig S. 1970. *Papers in Structural and Transformational Linguistics*. Netherlands: Springer.
Helasvuo, Marja-Liisa & Kyröläinen, Aki-Juhani. 2016. Choosing between zero and pronominal subject: modeling subject expression in the 1st person singular in Finnish conversation. *Corpus Linguistics and Linguistic Theory* 12:2, 263–299.
Ivaska, Ilmari. 2015. *Edistyneen oppijan suomen konstruktiopiirteitä korpusvetoisesti: avainrakenneanalyysi*. Turku: Turun yliopisto.
Jantunen, Jarmo Harri. 2012. Korpusvetoinen tekstilajianalyysi: sanalistat ja genreavainsanat. Julkaisussa: Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Lounela, Mikko & Tiililä, Ulla (toim.) *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Kotuksen julkaisuja 169. Helsinki: Gaudeamus, 360–371.

Krestel, Ralf & Bhaskar, Mehta. 2010. Learning the importance of latent topics to discover highly influential news items. Julkaisussa: Dillmann, Rüdiger, Beyrer, Jürgen, Hanebeck, Uwe D. & Schultz, Tanja (toim.) *Advances in Artificial Intelligence*. Berlin: Springer, 211–218.
Kyröläinen, Aki-Juhani. 2013. *Reflexive Space: A Constructionist Model of the Russian Reflexive Marker*. Annales Universitatis Turkuensis B 363. Turku: Turun yliopisto.
Lagus, Krista, Pantzar, Mika, Ruckenstein, Minna & Ylisiurua, Marjoriikka. 2016. Suomi24 – Muodonantoa aineistolle. Kuluttajatutkimuskeskus, Valtiotieteellisen tiedekunnan julkaisuja 2016:10. Helsinki: Helsingin yliopisto. Saatavissa: https://blogs.helsinki.fi/citizenmindscapes/files/2016/05/257383_HY_VALT_suomi24_muodonantoa_aineistolle.pdf.
Laippala, Veronika, Luotolahti, M. Juhani, Kyröläinen, Aki-Juhani, Salakoski, Tapio & Ginter, Filip. 2017. Creating register sub-corpora for the Finnish Internet Parsebank. Julkaisussa: Tiedemann, Jörg & Tahmasebi, Nina (toim.) *Proceedings of the 21st Nordic Conference on Computational Linguistics* (NoDaLiDa). Gothenburg: Linköping University Electronic Press, 152–161.
Laippala, Veronika, Kyröläinen, Aki-Juhani, Kanerva, Jenna & Ginter, Filip. 2018. Dependency profiles in the large-scale analysis of discourse connectives. *Corpus Linguistics and Linguistic Theory*. Saatavissa: https://doi.org/10.1515/cllt-2017-0031.
Lehti, Lotta, Luodonpää-Manni, Milla, Jantunen, Jarmo Harri, Kyröläinen, Aki-Juhani, Vesanto, Aleksi & Laippala, Veronika. Tulossa 2020. Commenting on poverty online: A corpus-assisted discourse study of Suomi24 forum. SKY Journal of Linguistics 33.
Michel, Jean-Baptiste, Shen, Yuan Kui, Aiden, Aviva Presser, Veres, Adrian, Gray, Matthew K., The Google Books Team, Pickett, Joseph P., Hoiberg, Dale, Clancy, Dan, Norvig, Peter, Orwant, Jon, Pinker, Steven, Nowak, Martin A. & Lieberman Aiden, Erez. 2011. Quantitative analysis of culture using millions of digitized books. *Science* 6014, 176–182.
Moisl, Hermann. 2015. *Cluster Analysis for Corpus Linguistics*. Berlin: De Gruyter Mouton.
Pagel, Mark, Atkinson, Quentin D. & Meade, Andrew. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* 7163, 717–720.
Quinn, Kevin, Monroe, Burt L., Colaresi, Michael, Crespin, Michael H. & Radev, Dragomir R. 2010. How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science* 54:1, 209–228.
Roberts, Margaret E., Stewart, Brandon M., Tingley, Dustin & Airoldi, Edoardo M. 2013. The structural topic model and applied social science. Paper presented at the conference ”Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation”, 1–20.
Roberts, Margaret E., Stewart, Brandon M. & Tingley, Dustin. 2016. Navigating the local modes of big data: The case of topic models. Julkaisussa: Alvarez, Michael R. (toim.) *Computational Social Science: Discovery and Prediction*. New York (NY): Cambridge University Press, 51–97.
Roemer, Ute & Wulff, Stefanie. 2008. Applying corpus methods to written academic texts: Explorations of MICUSP. *Journal of Writing Research* 2:2, 99–127.
Saukkonen, Pauli. 1977. *Nykysuomen saneiston yleisyystilastoa saneenloppuisessa aakkosjärjestyksessä*. Oulu: Oulun yliopisto.

Stubbs, Michael. 1995. Collocations and semantic profiles. On the cause of the trouble with quantitative studies. *Functions of Language* 2, 23–55.

Tagliamonte, Sali & Baayen, R. Harald. 2012. Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24:2, 135–178.

Tieteen termipankki. 2020. [verkkoaineisto]. [viitattu 28.4.2020]. Saatavissa: https://tieteentermipankki.fi.

Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R*. (2. p.). Boca Raton: Chapman & Hall/CRC Press.

Yule, Udny C. 1944. *The Statistical Study of Literary Vocabulary*. Cambridge: Cambridge University Press.

Zipf, George Kingsley. 1932. *Selective Studies and the Principle of Relative Frequency in Language*. Cambridge: Harvard University Press.

# Luku 11 Historiallinen korpuslingvistiikka

*Turo Vartiainen*
https://orcid.org/0000-0002-4760-750X
*Tanja Säily*
https://orcid.org/0000-0003-4407-8929

**Mikä?**

Historiallinen korpuslingvistiikka tarjoaa menetelmän eri kielten historian ja muutoksen tutkimiseen laajojen, systemaattisesti koottujen elektronisten tekstikokoelmien avulla. Menetelmässä yhdistyvät laadullinen ja määrällinen näkökulma, ja siinä korostuu tutkittavan aineiston tarkka tuntemus. Historiallisen korpuslingvistiikan menetelmin voidaan saada uutta tietoa paitsi kielestä järjestelmänä myös ihmisyhteisöjen kielikäytäntöjen vaihtelusta ja muutoksesta.

**Katso myös:**

Luku 7 Laadullinen aineistopohjainen kielentutkimus

Luku 9 Korpusaineistot

Luku 10 Määrällinen korpuslingvistiikka

Luku 15 Historiallinen ja vertaileva kielentutkimus

# 1. Johdanto

Kielen varhaisvaiheiden tutkimus on lähtökohtaisesti aina aineistopohjaista. Oli kyseessä sitten yksittäisen kielen muutos ajassa tai kokonaisen kielikunnan varhaisimman oletetun vaiheen, protokielen, rekonstruktio, tutkimus pohjautuu aina olemassa oleviin, usein sattumalta nykyaikaan saakka säilyneisiin teksteihin. Historiallinen korpuslingvistiikka on menetelmä, jonka avulla voidaan systemaattisesti tutkia eri kielten varhaisempia vaiheita monesta eri näkökulmasta. Kyseessä on empiirinen metodologia, jossa tutkittavan kieliaineiston niukkuus voi toisinaan asettaa tutkimukselle tiettyjä rajoituksia ja haasteita mutta tarjoaa samalla mahdollisuuden tarkastella paitsi hyvin yksityiskohtaisia kielioppiin, kielenmuutokseen ja kielen käyttöön liittyviä kysymyksiä myös abstraktimpia, perustavanlaatuisia kysymyksiä kielestä dynaamisena, ajassa muuttuvana ilmiönä. Yksityiskohtaisemmat, mikrotason tutkimuskysymykset, jotka kiinnostavat historiallisia korpuslingvistejä, voivat koskea esimerkiksi kielen taivutusjärjestelmän, apuverbirakenteiden tai sanajärjestyksen vaihtelua ja muutosta. Laajempia, makrotason teemoja, joihin on viime aikoina alettu kiinnittää yhä tarkempaa huomiota, ovat esimerkiksi seuraavat kysymykset:

1) Mitkä kielensisäiset tekijät edistävät kielenmuutosta ja millä tavoin?
2) Mitkä kielenulkoiset tekijät, kuten eri kieltä puhuvien ihmisten väliset kontaktit ja kieliyhteisöjen rakenne, vaikuttavat kielenmuutokseen?
3) Millaiset ihmiset tai ihmisryhmät omaksuvat erityisen nopeasti uusia tapoja käyttää kieltä ja tällä tavoin edesauttavat kielenmuutoksen leviämistä? Ketkä ovat konservatiivisia kielenkäyttäjiä?

Kuten näistä teemoista käy ilmi, suuri osa historiallisista korpuslingvisteistä on kiinnostunut nimenomaan siitä, millä tavoin ja missä olosuhteissa kieli muuttuu ajan saatossa. Tutkimuksen näkökulma on siis usein **diakroninen**, mutta kielten varhaisempia vaiheita on mahdollista

tarkastella myös **synkronisesta** näkökulmasta, ikään kuin läpileikkauksena tiettynä aikakautena puhutusta kielestä. Oli näkökulma kumpi hyvänsä, tutkija joutuu työssään ottamaan huomioon monenlaisia niin aineiston edustavuuteen kuin korpuslingvistiseen menetelmäänkin liittyviä asioita, jotka vaikuttavat siihen, millaisia kysymyksiä korpusaineiston avulla on ensinnäkin mahdollista tutkia, sekä siihen, millaisia johtopäätöksiä tulosten pohjalta voidaan esittää. Tutkijan vastuu on erityisen suuri silloin, kun käytössä on vain niukka ja tyylikirjoltaan suppea tai epäyhtenäinen aineisto, jonka perusteella on hankala tarkastella kielenmuutokseen liittyviä yleisluontoisia hypoteeseja (ks. luku 2).

Yksi historiallisen korpuslingvistiikan suurimmista eduista on sen tarjoama mahdollisuus kielen ilmiöiden yleisyyden eli frekvenssin tarkasteluun. Frekvenssitietojen avulla tutkija voi tehdä tarkkoja havaintoja ajassa vähitellen tapahtuvista muutoksista sekä tarkastella sellaisia kielen ilmiöitä, jotka eivät ilmene kategorisina joko–tai-jakoina vaan pikemminkin kielenkäytössä esiintyvinä tendensseinä, joiden ajallinen muutos voi lopulta johtaa koko kielioppijärjestelmän muuttumiseen. Koska korpusaineistossa on aina myös sattumanvaraista vaihtelua, tutkijat hyödyntävät tällaisten kielenmuutosten tutkimuksessa usein tilastollisia menetelmiä, joiden avulla voidaan esittää todennäköisyys sille, onko esimerkiksi kahtena eri aikakautena havaittavassa kielenmuutoksessa kyse todellisesta kielioppijärjestelmän muutoksesta vai onko kyse vain kielenkäyttöön luonnollisesti kuuluvasta vaihtelusta. Esittelemme joitain historiallisessa korpustutkimuksessa yleisesti käytettyjä tilastollisia menetelmiä myöhemmin tässä artikkelissa. Tässä yhteydessä on syytä todeta, että tilastollisten menetelmien käyttö ei historiallisessa korpuslingvistiikassa ole pakollinen osa tutkimusta. On mahdollista, että kielenmuutos etenee ajassa niin selkeästi, että tulokset ikään kuin ”puhuvat puolestaan”. Tilastollisia menetelmiä ei kuitenkaan ole syytä karttaa tai pelätä: tutkijoiden käytössä on nykyään monia helppokäyttöisiä tilastollisia testejä, jotka voivat tarjota aineistosta arvokasta lisätietoa.

Monet historiallista korpuslingvistiikkaa koskevista eduista, haasteista ja rajoitteista koskevat korpuslingvistiikkaa yleisemminkin (ks. Korpusaineistot tässä kirjassa). Nykykielen tutkimuksesta poiketen kielen varhaisempien vaiheiden tutkijat eivät kuitenkaan voi täydentää

korpusaineiston kautta saamaansa käsitystä kielen rakenteesta ja sen muutoksesta konsultoimalla kielen puhujia. Tätä voidaan pitää sekä etuna että haittana: yhtäältä on selvää, että kielenpuhujien intuitiot ja näkemykset omasta kielestään voivat tarjota arvokasta tietoa. Toisaalta on myös selvää, että puhujien intuitiot eivät aina pidä paikkaansa ja että heidän käsityksensä niin omasta kuin muidenkin ihmisten kielenkäytöstä voivat olla virheellisiä. Tästä näkökulmasta historiallinen korpuslingvistiikka näyttäytyy erityisen vahvasti aineistopohjaisena menetelmänä. Menetelmän aineistolähtöisyys ei kuitenkaan tarkoita sitä, etteivätkö historialliset korpuslingvistit voisi selittää tuloksiaan esimerkiksi kognitiivisesta tai sosiolingvistisestä näkökulmasta: vaikka tutkijoilla ei ole pääsyä menneinä aikoina eläneiden ihmisten ajatuksiin, psykologisiin prosesseihin ja käyttäytymismalleihin (esimerkiksi psykologisten testien, kielitajun, kyselytutkimusten tai havainnoinnin avulla), on oletettavaa, että esimerkiksi ihmisen aivotoiminta on biologisesta näkökulmasta ollut samanlaista koko kirjoitetun, n. 5 000 vuoden päähän yltävän historian ajan. Myös ihmisyhteisöjen ja ihmisten sosiaalisten roolien tarkka tutkimus ja ymmärtäminen tarjoavat mahdollisuuden selittää varhaisempien kieliyhteisöjen kielenkäyttöä ja sitä kautta kielenmuutoksen ulkoisia tekijöitä.

Historiallisen korpusaineiston avulla voidaan myös tarkastella ihmisen kognitioon ja kielenmuutokseen liittyviä yleisiä prosesseja. Yksi varsin paljon tutkittu esimerkki tällaisesta prosessista on merkitysten ns. **subjektifikaatio**, mikä tarkoittaa sitä, että ihmisillä on tapana alkaa käyttää sanoja ja rakenteita yhä enemmän omien mielipiteidensä, asenteidensa ja olettamustensa ilmaisuun – merkitykset ikään kuin hivuttautuvat lähemmäksi kielenpuhujaa, kielenkäytön subjektia (Benveniste 1971 [1958]: *le sujet de l'énonciation*). Esimerkiksi englannin kielessä yleisesti käytettyä *be going to* -rakennetta käytettiin alun perin ilmaisemaan vain fyysistä liikettä (esim. *She's going to the market*), kun taas nykyenglannissa rakennetta käytetään yleisemmin ilmaisemaan asioita, joita kielenpuhuja joko olettaa tapahtuvan tulevaisuudessa tai joiden tapahtumisesta hän on epävarma (esim. *It's going to rain; Are you going to come to the party?*). Vaikka tällaisia havaintoja kielenmuutoksesta ja sen luonteesta on mahdollista tehdä myös ilman korpusaineistoja, korpuslingvistiikka

tarjoaa monipuolisen menetelmän, jonka avulla voidaan hyvinkin tarkasti ajoittaa kielessä tapahtuvia muutoksia sekä etsiä niitä kielenkäytön konteksteja, jotka ovat olleet erityisen suotuisia muutoksen etenemiselle.

On syytä painottaa, että historiallinen korpuslingvistiikka on menetelmänä täysin riippumaton tutkittavasta kielestä – mitä tahansa kieltä, josta on koottu historiallinen korpus, on mahdollista tutkia historiallisen korpuslingvistiikan menetelmin, eikä ole lainkaan epätavallista, että tutkija itse koostaa korpuksen tutkimuksen kohteena olevasta kielestä. Koska tämän artikkelin kirjoittajien tausta on kuitenkin englannin historiallisessa kielitieteessä ja koska menetelmä on hyvin yleinen englannin kielen tutkimuksessa, jossa sillä on myös erityisen pitkä historia, artikkelissa käytetyt esimerkit ovat pääasiassa englannin kielestä.

## 2. Metodin historiaa ja taustaa

Historiallisen korpuslingvistiikan voidaan yleisesti katsoa pohjautuvan aineistoperäiseen kielentutkimukseen, jota on tehty vuosisatojen ajan. Uuden ajan Länsi-Euroopassa tutkijoiden mielenkiinnon kohteina olivat etenkin klassiset kielet, latina ja klassinen kreikka, joiden pohjalta alettiin jo 1800-luvulla koostaa kattavia tekstikokoelmia. Näistä kokoelmista on niiden ensi vaiheista lähtien käytetty nimitystä **korpus**, mutta on syytä painottaa, että *korpus*-termiä ei tuolloin käytetty sanan nykyisessä, teknisessä merkityksessä. Esimerkiksi Corpus Inscriptionum Latinarum -projektin pyrkimyksenä on ollut koota kaikki Rooman valtakunnan alueella laaditut latinankieliset piirtokirjoitukset yhteen kriittiseen editioon riippumatta niiden kirjoitusajasta, tarkoituksesta tai kirjoituksen alkuperästä. Nykymerkityksessään *korpuksella* tarkoitetaan kuitenkin tyypillisesti tarkoin harkittujen periaatteiden mukaisesti koottua digitaalista kieliaineistoa: tietokoneluettavaa tekstikokoelmaa, joka sisältää ennalta määritellyssä ja tarkoin harkitussa suhteessa erityyppisiä tekstejä, joiden voidaan katsoa riittävässä laajuudessa edustavan kieltä sen eri käyttöfunktioissa ja rekistereissä (McEnery & Hardie 2012). Ajatuksena on

se, että ihmiset käyttävät kieltä hyvin eri tavoin erilaisissa konteksteissa, ja jos kielitieteellinen korpus sisältää liian suppean valikoiman tekstejä – vaikkapa ainoastaan sanomalehtiartikkeleita –, jää tutkimuksen tavoittamattomiin kaikki se kielen käyttöön liittyvä vaihtelu, joka nousee esiin esimerkiksi epämuodollisissa keskusteluissa, sähköposteissa, blogikirjoituksissa tai kaunokirjallisuudessa. On toki mahdollista, että korpus sisältää tekstejä vain yhdestä genrestä (ks. alla). Tällöin tutkijan on kuitenkin varottava esittämästä liian yleisiä väitteitä kielestä aineistonsa perusteella: kielenmuutokset voivat näkyä eri genreissä eri aikoina, ja jotkin kielen rakenteet voivat olla hyvin yleisiä yhdessä genressä ja äärimmäisen harvinaisia toisessa. Esimerkiksi yksikön ensimmäisen ja toisen persoonan pronominit ovat erittäin yleisiä epämuodollisissa keskusteluissa, mutta niitä ei käytetä juuri lainkaan akateemisissa artikkeleissa.

Oli kyseessä sitten useammasta genrestä koostuva **yleiskorpus** tai vain yhden genren käsittävä korpus, korpuslingvistien tavoitteena on aina tehdä korpuksista mahdollisimman **edustavia**, niin että niiden voidaan ajatella riittävässä määrin kuvaavan kieltä sellaisena kuin se tietyn kieliyhteisön moninaisissa kielenkäyttötilanteissa esiintyy (ks. Korpusaineistot tk.). Toinen, korpuksen edustavuuteen liittyvä ja aivan yhtä tärkeä pyrkimys on saavuttaa **tasapaino** korpuksessa esiintyvien tekstien välillä. Yleiskorpuksessa voidaan esimerkiksi pyrkiä siihen, että eri tekstilajeja, kuten akateemisia artikkeleita, aikakauslehtiä, blogitekstejä ja epämuodollista keskustelua, on korpuksessa suhteellisesti yhtä paljon. Korpuksen tasapainoa voidaan myös parantaa ottamalla huomioon vaikkapa tekstien kirjoittajien sukupuoli ja ikä, jolloin korpusaineiston avulla voidaan tutkia myös sosiolingvistisiä kysymyksiä, kuten miesten ja naisten kielenkäyttöön liittyviä eroja. Korpuksen edustavuuteen ja tasapainoon liittyvät kysymykset koskevat niin synkronisia kuin diakronisia korpuksia, mutta historialliseen korpuslingvistiikkaan ja varsinkin diakroniseen tutkimukseen olennaisesti kuuluva aineistojen välinen vertailu korostaa entisestään näiden metodologisten perusperiaatteiden tärkeyttä.

Vaatimukset korpuksen edustavuudesta ja tasapainosta otettiin huomioon jo 1960-luvun alussa, kun ensimmäistä nykyaikaista

amerikanenglannin korpusta alettiin koota Brownin yliopistossa Yhdysvalloissa. Vuonna 1964 julkaistu ja yhä tutkijoiden käyttämä Brown Corpus[1] koostuu sanomalehtiteksteistä sekä kauno- ja tietokirjallisuudesta, ja sen sisältämät tekstit on luokiteltu yksityiskohtaisesti alakategorioihin: korpus sisältää n. 2 000 sanan otteita esimerkiksi scifi- ja seikkailukirjoista, elämäkerroista ja sanomalehtien pääkirjoituksista (ks. Francis ja Kučera 1979; ks. myös Korpusaineistot tk.). Kiinnostus kielen historiaan ja sen muutokseen johti 1990-luvun alussa uuden, samoilla periaatteilla toteutetun korpuksen kokoamiseen. Freiburg-Brown- eli Frown-korpusta[2] voidaan käyttää kolmekymmentä vuotta vanhemman Brown-korpuksen vertailukorpuksena. Sittemmin Brown-korpusperhettä on täydennetty 1960- ja 1990-luvun brittienglannin korpuksilla (Lancaster-Oslo-Bergen Corpus eli LOB[3]; Freiburg-LOB Corpus eli FLOB[4]) sekä kahdella 1930-luvun britti- ja amerikanenglannin korpuksella (BLOB-1931[5], The 1930s Brown Corpus eli B-Brown[6]). Koska Brown-perheen korpukset vastaavat niin hyvin toisiaan rakenteellisesti, ne tarjoavat erinomaisen mahdollisuuden tarkastella kielenmuutosta noin 60 vuoden aikavälillä sekä britti- että amerikanenglannissa.

Brown- ja Frown-korpukset noudattavat vaatimuksia korpuksen tasapainosta ja edustavuudesta siinä määrin kuin huoliteltu yleiskieli voi ylipäätään antaa tarkan kuvan kielestä, sen vaihtelusta ja muutoksesta. Näitä korpuksen kokoamista koskevia yleisiä periaatteita on kuitenkin sitä hankalampi noudattaa, mitä kauemmas taaksepäin kielen historiassa menemme: vanhoja tekstejä ei yksinkertaisesti ole aina säilynyt riittävästi, jotta valinnanvaraa olisi. Varhaisten kielivaiheiden tutkijat joutuvatkin usein työskentelemään sellaisen aineiston parissa, joka on säilynyt osin sattumalta ja osin taas siksi, että teksti on syystä tai toisesta joko koettu yhteiskunnallisesti tärkeäksi ja säilyttämisen arvoiseksi tai tuotettu olosuhteissa, jotka ovat jostain syystä edesauttaneet sen säilymistä. Valtaapitävien tekstit ovat kautta historian säilyneet huomattavasti todennäköisemmin kuin tavallisen kansan kirjoitukset, ja tämä näkyy myös monen historiallisen korpuksen tekstivalikoimassa: lakitekstit, luostarisäännöt ja raamatunkäännökset ovat yleensä varsin hyvin edustettuina, kun taas yksityishenkilöiden kirjoittamat kirjeet, päiväkirjat tai muut arkielämää koskevat tekstit ovat hyvin harvinaisia.

Myös kielen murrevaihtelu voi asettaa tutkimukselle omat haasteensa. Esimerkiksi valtaosa muinaisenglannin ajalta (n. 700–1150 jaa.) säilyneistä teksteistä on kirjoitettu Etelä-Englannissa sijainneen Wessexin kuningaskunnan murteella. Koska muinaisenglanniksi kirjoitettuja tekstejä on säilynyt nykypäivään saakka vain noin kolme miljoonaa sanaa, on tämä materiaali otettu kokonaisuudessaan osaksi eräitä englannin kielen historiallisia korpuksia. Tällöin ongelmaksi muodostuu se, että valtaosa englannin kielen seuraavan vaiheen, keskienglannin (n. 1150–1500 jaa.), ajalta säilyneistä teksteistä on peräisin Wessexiä pohjoisemmilta Itä- ja Länsi-Midlandsin alueilta. Tällaisia korpuksia työssään käyttävä kielenmuutoksen tutkija joutuukin pitämään mielessään, että aineistossa havaittava muutos voi selittyä sillä, että eri murteiden suhteellinen osuus ei ole kaikkina korpuksen edustamina aikakausina tasapainossa. Tilanne voi pahimmillaan vastata sitä, että suomen kielen lähihistoriaa tutkittaisiin vertailemalla 1800-luvun Kuopion murretta 2000-luvun Turun murteeseen ja tämän perusteella tehtäisiin yleisluontoisia, koko suomen kieltä koskevia johtopäätöksiä.

Yksi varhaisimmista, ja samalla kunnianhimoisimmista, historiallisista korpuksista on Helsinki Corpus of English Texts (HC)[7], jonka kokoaminen aloitettiin 1980-luvulla Helsingin yliopistossa professori Matti Rissasen johdolla. Helsinki Corpus julkaistiin vuonna 1991, ja sen pohjalta on sittemmin kirjoitettu satoja tieteellisiä artikkeleita, jotka valaisevat englannin kielessä tapahtunutta muutosta muinaisenglannista 1700-luvulle. Korpuksen vahvuutena on sen kokoamistyössä noudatettu huolellisuus: eri aikakausien tekstit on luokiteltu yksityiskohtaisiin kategorioihin, ja pyrkimyksenä on ollut esittää englannin kielen yli tuhatvuotinen jatkumo niin, ettei kokonaiskuva liiaksi hämärtyisi murrevaihtelun tai tekstilajeissa tapahtuvien muutosten vuoksi. Helsinki Corpus koottiin aikana, jolloin yhteiskunnan tietokoneistuminen oli vasta alkanut. Tämä näkyy omalta osaltaan myös korpuksen koossa: korpus käsittää kaikkinensa n. 1,5 miljoonaa sanaa, mikä riittää vain suhteellisen yleisten kielioppi-ilmiöiden ja sanastonmuutoksen tutkimiseen. Toisaalta se, mikä koossa hävitään, voitetaan tarkkuudessa. Viime vuosina julkaistut ns. ”megakorpukset”, kuten yli 400 miljoonaa sanaa vuosilta 1810–2009 käsittävä Corpus of Historical American English

(COHA)[8], ovat kokoamisperiaatteiltaan melko suurpiirteisiä: tekstien kategorisointi on toteutettu karkeammin, ja koska korpuksen suuren koon vuoksi ei tutkijoilla ole ollut mahdollisuutta tarkastaa tekstejä manuaalisesti, on korpuksessa jonkin verran niin tekstien ajoitusta, kirjoitusmuotoja kuin kielen varianttiakin koskevia virheitä.

Riippumatta käytetystä korpuksesta kielentutkijan vastuulla on aina tuntea aineistonsa mahdollisimman tarkasti, niin että aineiston epätasapaino tai eri tekstilajien konventioissa tapahtuvat muutokset eivät johda vääriin tulkintoihin. Esimerkiksi kaikki historialliset tekstilajit, kuten pyhimyselämäkerrat tai erilaiset pamfletit, eivät sellaisinaan ole säilyneet nykypäivään, ja tämä on hyvä pitää mielessä korpusaineistoa tutkittaessa (ks. myös Heikkinen, Voutilainen, Lauerma, Tiililä & Lounela 2012). Toisaalta voi syntyä uusia genrejä: esimerkiksi sanomalehdet kehittyivät pääasiassa 1700-luvulta lähtien, ja internetin yleistymisen myötä on viime vuosikymmeninä kehittynyt monia uusia tekstilajeja. Myös tekstilajeja koskevat käytänteet muuttuvat usein, ja tällä voi olla vaikutusta korpusaineistosta saatuihin tietoihin: esimerkiksi yksityiskirjeiden muodollisuus ja siihen liittyvät kielelliset piirteet sekä tieteellisen tekstin syntaktinen kompleksisuus ovat muuttuneet ajan saatossa. Kuitenkaan nämä esimerkiksi pronominien ja passiivirakenteiden frekvenssissä näkyvät muutokset eivät ole todiste kielenmuutoksesta vaan selittyvät yksittäisten tekstilajien esitystavoilla ja niihin liittyvillä muutoksilla.

Historiallisen korpuslingvistiikan aineistojen kehityksessä on nähtävissä useita eri tendenssejä. Toisaalta on saatavilla yhä suurempia korpuksia ja tietokantoja, kuten Early English Books Online (EEBO) ja Kansalliskirjaston lehtikokoelma, jotka sisältävät massiivisen määrän painettuja tekstejä ja joiden avulla voidaan tutkia entistä harvinaisempia muutoksia.[9] Toisaalta myös pienempiä korpuksia kootaan edelleen, mutta niiden annotointi on aiempaa kehittyneempää: puhutaan ns. ”rikkaasta datasta”, josta voi löytyä genre- ja sanaluokka-annotaation lisäksi vaikkapa kirjoittajien taustatietoja, tarkkoja kuvauksia tekstin asettelusta ja käytetyistä käsialoista tai kirjasintyypeistä sekä valokuvia alkuperäisteksteistä. Siinä missä varhaiset miljoonan sanan korpukset pyrkivät edustamaan kieltä yleensä ja sisälsivät useita genrejä, nykyisin pienet korpukset keskittyvät usein vain yhteen genreen, kuten kirjeisiin (esim.

Corpora of Early English Correspondence, CEEC)[10] tai jopa yhden kirjoittajan teksteihin. Tällainen korpus voidaan koota myös osana opinnäytetyötä, jossa voidaan sitten käyttää korpusta tutkimukseen (esim. Marttila 2014). Sekä isoa että rikasta dataa käytetään yhä enemmän digitaalisissa ihmistieteissä yleensä, ja samoja aineistoja voivat hyödyntää niin historioitsija, lingvisti kuin kirjallisuudentutkijakin.

# 3. Kuinka historiallista korpuslingvistiikkaa harjoitetaan

## 3.1. Tutkimuksen kulku

### 3.1.1. Aineiston hankkiminen ja käsittely

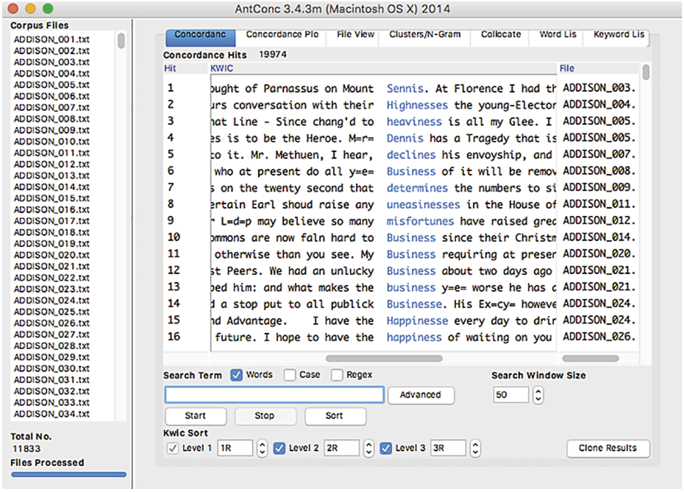
Historiallisessa korpuslingvistiikassa tutkimus aloitetaan tyypillisesti tekemällä aineistosta hakuja, joilla tavoitetaan se kielenaines, jonka muutosta halutaan tutkia. Tarvitaan siis sopiva korpus sekä jokin hakutyökalu. Historialliset korpukset voidaan lajitella kahteen luokkaan: Alun perin diakronisiksi kootuista korpuksista (esim. Helsinki Corpus ja Vanhan kirjasuomen korpus)[11] löytyy tekstejä enemmän tai vähemmän tasaisesti koko niiden kattamalta aikaperiodilta. Toisaalta on myös synkronisia korpuksia, jotka edustavat vain yhtä ajanhetkeä mutta joita on myöhemmin laajennettu kokoamalla vastaava korpus eri ajanhetkeltä. Jälkimmäistä tyyppiä edustaa aiemmin mainittu Brown-korpusperhe (ks. Korpusaineistot tk.).

Korpuksia on saatavilla monista eri lähteistä. Esimerkiksi Kielipankkiin[12] on koottu laaja kokoelma erikielisiä korpusaineistoja. Kielipankki on osa FIN-CLARIN-konsortiota, joka puolestaan kuuluu eurooppalaiseen CLARIN-organisaatioon, jonka tarkoituksena on kehittää ja tarjota kieliaineistoja sekä näiden aineistojen käsittelyyn sopivia työkaluja. Korpuksia voi etsiä myös metadatapalveluista. Historialliselle korpuslingvistiikalle ei vielä ole omaa palvelua, mutta yleisimpiin palveluihin lukeutuvat muiden muassa META-SHARE, CLARINin Language Resource

Inventory, Linguistic Data Consortiumin luettelo sekä englannin kielen osalta Corpus Resource Database (CoRD).[13] Korpukset eivät useinkaan ole täysin vapaasti saatavilla mm. tekijänoikeusongelmien takia, ja osa niistä on maksullisia. Yliopistojen kielten laitoksilla on tyypillisesti lisenssejä yleisimpien korpusaineistojen käyttöön, joten ennen lisenssimaksun maksamista kannattaa tarkistaa Kielipankin lisäksi oman yliopiston aineistokokoelmat.

Korpushakutyökalu tulee usein korpuksen mukana: Esimerkiksi COHA-korpusta voi käyttää sille räätälöidyllä web-käyttöliittymällä. Korpuksen kokoaja, Mark Davies, on laatinut muitakin historiallisia korpuksia, jotka ovat kaikki käytettävissä vastaavan nettikäyttöliittymän avulla (esim. espanjan- ja portugalinkieliset Corpus del Español ja Corpus do Português).[14] Suomen web-pohjaisista palveluista mainittakoon Kielipankin Korp[15] (ks. myös Korpusaineistot tk.). Korp-palveluun on sovitettu useita koti- ja ulkomaisia korpuksia, joista osa on diakronisia ja osa synkronisia. Kansainvälisesti yleinen web-pohjainen käyttöliittymä on CQPweb; esimerkiksi Lancasterin yliopiston CQPweb-palvelimelle on asennettu laaja kirjo etenkin englanninkielisiä korpuksia.[16]

Joitakin korpuksia voi ladata myös omalle koneelle ja käyttää erillisen konkordanssiohjelman avulla. Tällaisia ohjelmia ovat esimerkiksi maksullinen WordSmith Tools ja ilmainen AntConc (kuva 1), joiden käyttöön on internetissä hyvät ohjeet.[17] Konkordanssiohjelman perustoimintoihin kuuluu haku, jonka tulokset näytetään yleensä kuvan 1 tapaisena konkordanssilistauksena. Kukin hakutulos näytetään omalla rivillään, jonka keskellä on haettu sana, fraasi tai sanan osa, ja ympärillä konteksti, jossa hakutulos esiintyy. Ohjelmointitaitoiset voivat tehdä korpushakuja myös ohjelmointikielellä, vaikkapa R-ympäristössä.[18] Tämän lähestymistavan etuja on se, että tutkija ei ole sidottu valmiissa korpusohjelmissa oleviin toimintoihin, sekä se, että tutkimuksen voi periaatteessa helposti toistaa ajamalla tutkijan käyttämän skriptin.




Kuva 1. AntConc-ohjelmalla on haettu CEEC-korpuksesta englannin *-ness*-johtimella muodostettuja sanoja eri kirjoitusasuineen (esim. *-nes*, *-nis*). Osumien joukossa on paljon roskaa, kuten erisnimiä ja *-ne*-loppuisten sanojen taivutusmuotoja.

Kun alustavat haut on tehty, on hakutulokset käytävä läpi väärien osumien poistamista ja tulosten luokittelua varten (ks. Laadullinen aineistopohjainen kielentutkimus tk.). Osa yllä mainituista korpuskäyttöliittymistä tukee tällaista luokittelua, mutta usein on helpompaa siirtää tulokset erilliseen tiedostoon, jonka voi sitten avata Excelissä tai muussa taulukkolaskentaohjelmassa ja tehdä annotoinnin siellä. Monista korpusohjelmista tällainen siirto- tai export-toiminto löytyykin, mutta esimerkiksi COHA-käyttöliittymästä tiedot on siirrettävä selaimen kopiointitoimintoa käyttäen. Hakutulosten luokittelu ja siivoaminen on usein tutkimuksen aikaa vievin osa.

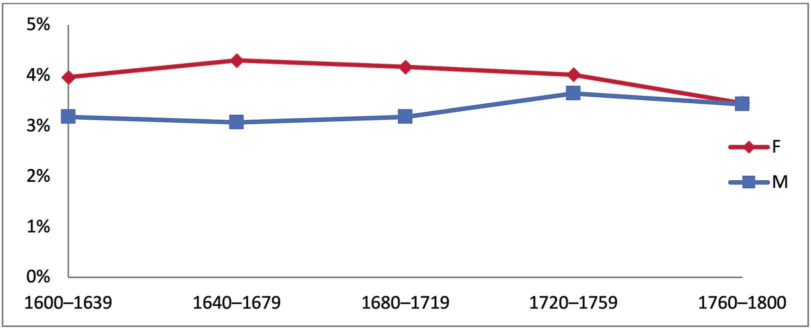
### 3.1.2. TILASTOLLINEN ANALYYSI JA VISUALISOINTI

Luokitelluista hakutuloksista voidaan siirtyä tekemään tilastollista analyysia. Tämä onnistuu valitusta menetelmästä riippuen joidenkin korpustyökalujen sisällä, taulukkolaskentaohjelmilla, erityisillä tilasto-ohjelmilla, kuten SPSS:llä[19], tai itse ohjelmoiden (viimeiseen vaihtoehtoon hyvä opaskirja on Gries 2013). Yleisin kvantitatiivinen menetelmä historiallisessa

korpuslingvistiikassa on laskea tutkittavan ilmiön **normalisoitu frekvenssi** aikakausittain ja vertailla frekvenssejä. Näin saadaan tietoa ilmiön esiintymistiheyden muutoksesta ajassa. Normalisoitu frekvenssi lasketaan jakamalla esiintymien lukumäärä koko aikakauden sanamäärällä; tämä voidaan kertoa vaikkapa kymmenellä tuhannella, jolloin saadaan ilmiön esiintyvyys kymmentä tuhatta sanaa kohti. Aikakausien sisällä voidaan tarkastella eri alakorpuksia, kuten genrejä tai sosiaalisia kategorioita, jolloin näiden normalisoitu frekvenssi lasketaan jakamalla alakorpuksen esiintymien lukumäärä alakorpuksen sanamäärällä. Esimerkiksi kuvassa 2 tarkastellaan englannin *I*-persoonapronominin normalisoidun frekvenssin muutosta ajassa naisten ja miesten alakorpuksissa. Kuvasta käy ilmi, että *I*-pronominin osuus on suurempi naisten kuin miesten kirjeissä kaikilla paitsi viimeisellä aikakaudella (ks. tietolaatikko 1).

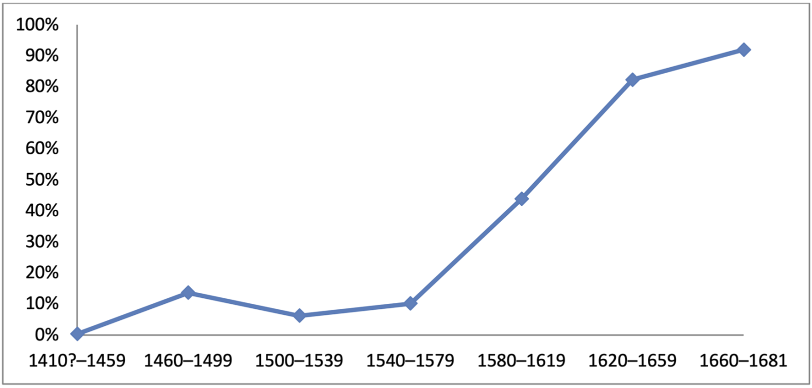
**Tietolaatikko 1.**

- Historiallinen sosiolingvistiikka tutkii historiallisen korpuslingvistiikan keinoin muun muassa sitä, miten kielenkäyttö vaihtelee eri yhteiskuntaryhmien kesken ja mitkä ryhmät johtavat kielenmuutosta (Nevalainen & Raumolin-Brunberg 2003).
- Vartiainen, Säily ja Hakala (2013) tutkivat englannin ensimmäisen ja toisen persoonan pronominien vaihtelua ja muutosta CEEC-korpuksessa. He havaitsivat, että naiset käyttivät näitä pronomineja keskimäärin enemmän kuin miehet. Tulos liittyy sukupuolittuneisiin kirjoitustyyleihin, joista on englannin kielestä evidenssiä vuosisatojen ajalta: miesten tyyli on usein informatiivisempi, naisten taas keskittyy enemmän vuorovaikutukseen.
- On kuitenkin huomattava, että tyylit eivät pysy täysin muuttumattomina: esimerkiksi vaikuttaa siltä, että 1700-luvulla naisten ja miesten kirjoitustyylit lähenivät toisiaan. Lisäksi erot tulevat esiin vain keskiarvoissa, ja yksilöiden välillä on paljon vaihtelua (ks. kuva 4). Lisäksi kielenkäyttöön vaikuttavat tarkemmat sosiaaliset roolit: esimerkiksi puolisoiden välisissä kirjeissä miehet saattoivat käyttää *I*-pronominia yhtä paljon kuin naiset (Vartiainen, Säily & Hakala 2013).




Kuva 2. Englannin *I*-persoonapronominin normalisoitu frekvenssi (ilmaistuna prosenttiosuutena kaikista sanoista) 1600-luvulta 1800-luvulle CEEC-korpuksessa. Aikakaudet on jaettu alakorpuksiin kirjoittajan sukupuolen perusteella (F = naiset, M = miehet).

Toinen normalisointimenetelmä otetaan käyttöön silloin, kun on olemassa vaihtoehtoisia tapoja sanoa sama asia. Esimerkiksi englannin yksikön 3. persoonan preesensin verbipäätteestä on ollut olemassa kaksi eri varianttia, -*s* ja -*th* (esim. *has* vs. *hath*). Nykyään vanhempaan -*th*-muotoon törmää lähinnä enää uskonnollisissa teksteissä ja mahdollisesti joissakin englannin murteissa. Variantit muodostavat kielellisen muuttujan eli variaabelin, -*s*/-*th*. Kun halutaan tutkia -*s*:n yleistymistä ajassa, lasketaan sen esiintymien osuus koko variaabelista eli -*s*:n ja -*th*:n yhteenlasketuista esiintymistä kullakin aikakaudella (kuva 3). Huomataan, että viivadiagrammi muistuttaa muodoltaan hieman S-kirjainta: aluksi muutos etenee hitaasti, sitten tulee nopea vaihe, ja lopuksi muutoksen eteneminen hidastuu jälleen. Monet muutokset noudattavat tällaista S-käyrää, mutta jotkin saattavat tehdä jopa U-käännöksen.




Kuva 3. Englannin yksikön 3. persoonan preesensin verbipääte *-s*:n osuus variaabelista *-s*/*-th* CEEC-korpuksessa, 1410?–1681 (Nevalainen & Raumolin-Brunberg 2003).

Korpuksen jakaminen aikakausiin eli **periodisaatio** on oma taiteenlajinsa, jossa tasapainotellaan kuvauksen tarkkuuden ja datan määrän välillä. Usein korpuksen kokoajat ovat alustavasti jakaneet korpuksen tietynmittaisiin aikakausiin, joille he myös ilmoittavat korpuksen manuaalissa sanamäärät. Kaudet voivat olla mielivaltaisen mittaisia, tai ne voivat perustua esim. kielen yleisesti tunnettuihin vaiheisiin, historiallisiin aikakausiin tai sukupolven keskimääräiseen pituuteen. Viimeinen vaihtoehto on yleinen sosiolingvistisissä korpuksissa, kuten CEEC:ssä, joskin aineiston vähyyden vuoksi sukupolviin perustuvat 20-vuotiskaudet joudutaan toisinaan yhdistämään 40-vuotiskausiksi (ks. kuvat 2 ja 3). Helsinki Corpus puolestaan on jaettu englannin yleisen kehityksen mukaisesti muinaisenglantiin (–1150), keskienglantiin (1150–1500) ja varhaisuusenglantiin (1500–1710). Nämä kaudet on jaettu edelleen noin sadan vuoden mittaisiin jaksoihin, jotka ovat sopivan kokoisia tutkittaviksi. Toisaalta esimerkiksi Gries ja Hilpert (2012) ovat esittäneet, että periodisaation tulisi olla muutoskohtaista ja perustua kunkin muutoksen omiin vaiheisiin; he käyttävät vaiheiden tunnistamiseen tilastollista klusterointimenetelmää (ks. Määrällinen korpuslingvistiikka tk.).

Jos normalisoiduissa frekvensseissä havaitaan muutosta ajassa, herää heti kysymys, onko kyseessä todellinen kielenmuutos vai voisiko

se johtua satunnaisesta vaihtelusta tai aineiston epätasaisuudesta. Normalisoitujen frekvenssien vertailun tukena voidaan käyttää **hypoteesitestausta** (esim. Coolidge 2013; ks. myös Määrällinen korpuslingvistiikka tk.), joka kertoo, onko kahden aikakauden välinen ero tilastollisesti merkitsevä. **Nollahypoteesina** on, että ero on sattumaa, ja ero on merkitsevä vain, mikäli todennäköisyys (*p*-arvo) sille, että nollahypoteesi hylätään väärin perustein, on valittua **merkitsevyystasoa** alhaisempi (tasona käytetään usein viittä prosenttia, $p < 0{,}05$). Todennäköisyys lasketaan **merkitsevyystestillä**, joita on useita erilaisia. Historiallisessa korpuslingvistiikassa on perinteisesti käytetty $\chi^2$- eli khiin neliö -testiä, joka on helppo tehdä: tarvitsee tietää vain ilmiön esiintymien lukumäärät kummaltakin aikakaudelta sekä aikakausien kokonaissanamäärät, ja *p*-arvon voi katsoa taulukosta tai käyttää nettilaskuria.

Valitettavasti $\chi^2$-testi ja muut niin kutsutut sanasäkkitestit (engl. *bag-of-words tests*) eivät ota huomioon sanojen jakautumista korpuksessa, vaan ne olettavat, että sanat esiintyvät toisistaan riippumatta, mikä ei tietenkään pidä paikkaansa. Tämä tekee testeistä ylioptimistisia, eli ne pitävät eroa usein virheellisesti merkitsevänä. Oletetaan esimerkiksi, että korpuksessa on kaksi aikakautta, joissa kummassakin on 50 000 juoksevaa sanaa. Jos jokin sana esiintyy ensimmäisellä aikakaudella vaikkapa 25 kertaa ja toisella aikakaudella vain 5 kertaa, tällaiset testit pitävät eroa käytännössä aina merkitsevänä (taulukko 1). Testit eivät kuitenkaan ota huomioon sitä mahdollisuutta, että ensimmäisen aikakauden 25 esiintymää voivat olla kaikki samassa tekstissä, jolloin korkea frekvenssi on pelkkää sattumaa (Säily 2014, 49).

| | **Aikakausi 1** | **Aikakausi 2** |
|---|---|---|
| **Sana *X*** | 25 | 5 |
| **Jokin muu sana kuin *X*** | 49 975 | 49 995 |

Taulukko 1. Esimerkki khiin neliö -testissä käytettävästä kontingenssitaulusta, johon laitetaan sanamäärätiedot. Tilasto-ohjelman antama *p*-arvo on tässä tapauksessa 0,0005, eli $\chi^2$-testin mukaan on vain 0,05 %:n mahdollisuus, että nollahypoteesi hylätään väärin perustein.

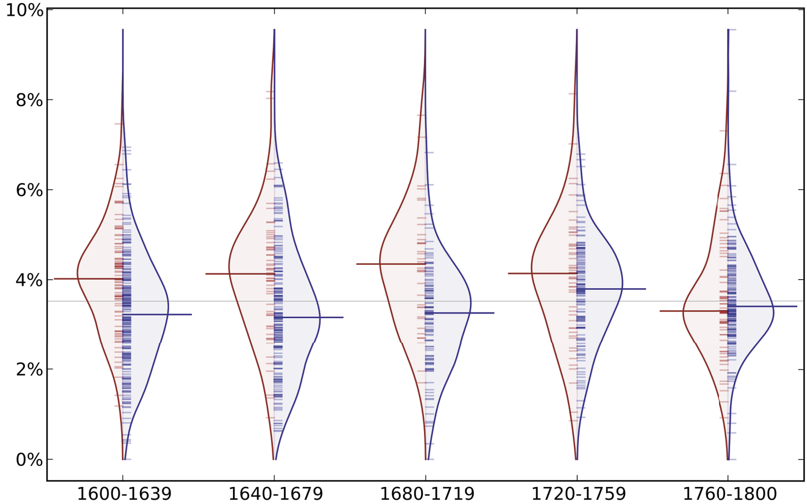
Näistä syistä korpuslingvistiikassa ollaankin siirtymässä ns. hajontatietoisiin testeihin (engl. *dispersion-aware tests*), joihin lukeutuvat mm. t-testi, Wilcoxonin järjestyssummatesti, Kendallin tau -testi ja uudelleenotantaan perustuvat testit (Säily 2014, 46). Näitä testejä pystytään käyttämään siten, että ne huomioivat sanojen jakautumisen korpuksessa, mutta siinä tapauksessa niiden käyttö on hankalampaa, sillä syötteeksi eivät riitä aikakausitason frekvenssit vaan on mentävä tekstitasolle: pitää tietää ilmiön frekvenssi kussakin korpuksen tekstissä sekä kunkin tekstin kokonaissanamäärä. Testeistä aloittelijalle helpoin ja luotettavin lienee Wilcoxonin järjestyssummatesti, toiselta nimeltään Mann-Whitneyn U-testi (esim. Metsämuuronen 2011, 386–390, 1102–1107), jonka pystyy tekemään Excelissä ja jolle on myös nettilaskuri. Tilastollisen merkitsevyyden osoittamiseen liittyy toki myös monimutkaistavia tekijöitä: mitä useampia hypoteeseja testataan, sitä suuremmalla todennäköisyydellä nollahypoteesi kumotaan väärin perustein. Tämän välttämiseksi voidaan käyttää erilaisia menetelmiä, jotka käytännössä tiukentavat merkitsevyystasoa (esim. Säily 2014, 49–51).

Historiallisessa korpuslingvistiikassa voidaan käyttää myös sofistikoituneempia tilastollisia menetelmiä, joiden avulla pystytään analysoimaan useiden ilmiöiden muutosta tai useita yhden ilmiön esiintyvyyteen vaikuttavia tekijöitä yhtä aikaa. Tällaista **monimuuttuja-analyysia** on käytetty esimerkiksi rakennevaihtelun, kuten englannin -*s*- ja *of*-genetiivien vaihtelun, tutkimiseen (Szmrecsanyi, Rosenbach, Bresnan & Wolk 2014) sekä genremuutoksen tutkimiseen Douglas Biberin kehittämää piirrepatteristoa käyttäen (Biber 1988; Biber & Finegan 1997). Biberin multidimensionaalinen menetelmä laskee kymmenien kielenpiirteiden frekvenssit teksteissä ja etsii niistä faktorianalyysin keinoin piirteet, joiden frekvenssit vaihtelevat samalla tavoin tekstien välillä. Näillä piirteillä tulkitaan olevan jokin kommunikatiivinen syy esiintyä yhdessä, ja eri aikakausia ja genrejä voidaan vertailla sijoittamalla ne piirteiden muodostamille rekisteriulottuvuuksille. Tämäntyyppinen tutkimus osoittaa, että kaikki muutos aineistossa ei ole kielenmuutosta, vaan se voi liittyä myös genrekonventioiden muutokseen. Esimerkiksi englannin historiassa on ollut käynnissä jo vuosisatoja kestänyt prosessi, jossa genret epämuodollistuvat (engl. *colloquialization*), jolloin niissä käytetään yhä

puhekielenomaisempia piirteitä, kuten persoonapronomineja sekä ihmisen ajatuksiin ja mielipiteisiin liittyviä verbejä (esim. *think*, *guess*).

Nykypäivän korpuslingvistiikassa yleiset avainsana-analyysi ja kollokaatioiden tutkimus (ks. Määrällinen korpuslingvistiikka tk.) ovat historiallisessa korpuslingvistiikassa olleet harvinaisempia, osittain aineiston vähäisyyden, osittain taas sanojen kirjoitusasujen vaihtelun takia. Ennen oikeinkirjoituksen standardisointia sanojen kirjoitusasuissa oli valtava määrä vaihtelua – esimerkiksi Parsed Corpus of Early English Correspondence[20] -kirjekorpuksessa (1410–1681) englannin kielen *tomorrow*-sanaa on kirjoitettu ainakin seuraavilla tavoilla: *to marrow, to moroe, to moroughe, to morow, to morowe, to morroughe, to morrow, to morrowe, to morue, to morwe, to-morow, to-morowe, to-morrow, to-morrowe, to-morw, to-morwe, tomorow, tomorowe, tomorrow, tomorrowe, too morrow, toomorrow*. Joistakin historiallisista korpuksista on nyttemmin tehty normalisoituja versioita, joissa tekstien kirjoitusasu on standardisoitu esim. VARD-ohjelmalla (Baron 2011); automaattinen standardisointi ei kuitenkaan pure harvinaisempiin kirjoitusasuihin, joten osa korpuksien sanoista jää edelleen standardisoimatta. Yksi harvoista historiallisista avainsana-analyyseista on Lijffijt, Säily & Nevalainen (2012); kollokaatioiden muutosta puolestaan ovat tutkineet mm. Baker, Brezina ja McEnery (2017), jotka keskittyvät 1800–1900-lukuihin, jolloin englannin kirjoittaminen oli etenkin painetuissa teksteissä jo melko yhtenäistä.

Tilastollisten menetelmien rinnalle tärkeiksi ovat nousemassa **tiedon visualisoinnin** menetelmät. Visualisointi on tiedon esittämistä tiivistetyssä muodossa; on osoitettu, että ihminen saa tietoa näköaistin kautta enemmän kuin kaikkien muiden aistien kautta yhteensä (Ware 2004). Yksinkertaisia, kuvien 2 ja 3 kaltaisia viivadiagrammeja voi hyvin tehdä vaikkapa Excelissä, mutta vaativampia visualisointeja varten tarvitaan jotain muuta, kuten R-ympäristössä toimiva ggplot2-kuvantamisjärjestelmä.[21] Historiallisessa korpuslingvistiikassa on nähtävissä suuntaus viivadiagrammeista periodin sisäistä vaihtelua ilmentäviin kuviin, joista yhtenä esimerkkinä mainittakoon *beanplot* eli papukaavio (kuva 4; Kampstra 2008).

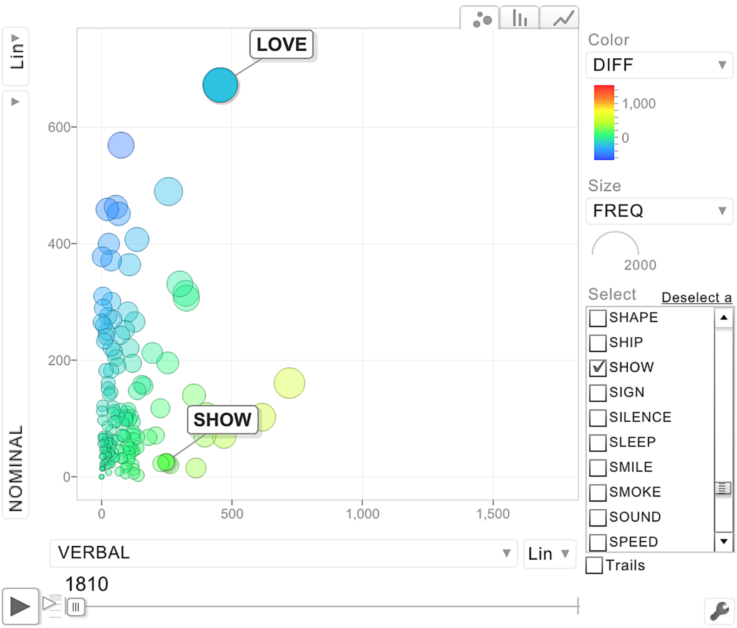



Kuva 4. Englannin I-persoonapronominin vaihtelu ja muutos CEEC-korpuksessa esitettynä *beanplot*-kaaviona (vrt. kuva 2). Naiset punaisella, miehet sinisellä. (Vartiainen, Säily & Hakala 2013.)

Kuva 4 pohjautuu samaan aineistoon kuin kuva 2, mutta se antaa paljon enemmän informaatiota. Paksut vaakaviivat kertovat *I*-pronominin mediaanifrekvenssin naisilla (vas.) ja miehillä (oik.) aikakausittain, kun taas kukin pienempi vaakaviiva ilmaisee pronominin normalisoidun frekvenssin yksittäisen henkilön kirjoittamissa teksteissä (mediaani on näistä havainnoista keskimmäinen). Näemme siis yhdellä silmäyksellä muutoksen yleiskuvan lisäksi sen, moneenko eri havaintoon mediaanit perustuvat: esimerkiksi keskimmäisellä kaudella naisten alakorpuksessa on suhteellisen vähän havaintoja, joten emme välttämättä luota sen mediaaniin yhtä paljon kuin vaikkapa ensimmäisen kauden. Wilcoxonin järjestyssummatestin mukaan naisten ja miesten välinen ero on tilastollisesti merkitsevä kahdella ensimmäisellä kaudella (Vartiainen, Säily & Hakala 2013, 240). Kaaviosta näkyy myös, miten havainnot jakautuvat alakorpuksien sisällä. Kunkin alakorpuksen havaintojen ympärille on piirretty jakauma (joka kaavion kehittäjän mielestä muistuttaa pavun palkoa), joka osoittaa, mihin suurin osa havainnoista keskittyy. Tämä auttaa huomaamaan aineiston epänormaaliudet: jos esimerkiksi jokin

osajoukko havainnoista käyttäisi pronominia huomattavan paljon ja jokin toinen taas vähän, jakaumassa näkyisi kaksi erillistä keskittymää.

Uusinta uutta korpuslingvistiikassa on interaktiivinen visualisointi, ja historiallinen korpuslingvistiikka on tässäkin ajan hermolla. Esimerkiksi Hilpert (2011) käyttää animoituja pistekaavioita (engl. *motion charts*), joissa kielenmuutos näkyy eri kielen ilmiöitä esittävien pisteiden liikkeenä kaaviossa. Pisteen koko kertoo sen normalisoidun frekvenssin. Kuvassa 5 näkyy joukko englannin sanoja, joita voidaan käyttää sekä substantiiveina (engl. *nominal*, y-akseli) että verbeinä (engl. *verbal*, x-akseli). Kuvasta ilmenee vuoden 1810 tilanne COHA-korpuksessa, mutta vasemman alakulman nuolta painamalla alkaa animaatio, joka osoittaa, miten sanojen käyttö muuttuu ajassa aina 2000-luvulle asti. Esimerkiksi *love*-sanaa käytetään aluksi lähinnä substantiivina, mutta




Kuva 5. Animoitu pistekaavio ambikategoristen sanojen muutoksesta COHA-korpuksessa (http://members.unine.ch/martin.hilpert/motion.html).

ajan myötä se muuttuu yhä ambikategorisemmaksi, eli sitä käytetään yhtä lailla sekä substantiivina (*What is love?*) että verbinä (*I love you*). Animoidut pistekaaviot auttavat tutkijaa prosessoimaan paljon tietoa tehokkaasti, ja niiden avulla on helppo saada yleiskuva muutoksesta sekä toisaalta seurata yksittäisten sanojen tai ilmiöiden muutosta.

On huomattava, että numerot tai kuvat itsessään eivät ole vielä tutkimustulos, vaan niiden tulkinnassa tarvitaan myös laadullista tutkimusta. Esimerkiksi pronominitutkimuksessamme (kuvat 2 ja 4, tietolaatikko 1) palasimme toistuvasti korpusteksteihin: luimme miesten ja naisten kirjoittamia kirjeitä saadaksemme selville, mistä sukupuoliero persoonapronominien käytössä johtui. Tarkastelimme, miten miehet ja naiset käyttivät pronomineja ja miten tämä liittyi heidän sosiaalisiin rooleihinsa isinä, poikina, aviomiehinä, vaimoina ja äiteinä. Havaitsimme muun muassa, että vaikka 1600-luvun miehet ja naiset käyttivät puolisoiden välisissä kirjeissä *I*-pronominia keskimäärin yhtä paljon, he tekivät pronominilla eri asioita: aviomiehet kertoivat tyypillisesti siitä, mitä he olivat tehneet (esim. *I met my old friend*), kun taas vaimot kertoivat omista ajatuksistaan ja tunteistaan (esim. *I think*); toisaalta vaimot käyttivät enemmän *you*-pronominia, koska hekin keskittyivät aviomiehen tekemisiin. Lisäksi kytkimme löydökset historiallisiin konteksteihinsa, kuten siihen, että aviopuolisoiden välinen suhde oli tuon ajan Britanniassa epätasa-arvoinen ja mies oli suhteen hallitseva osapuoli. Seuraava luku tarjoaa lisää esimerkkejä siitä, mitä historiallisen korpuslingvistiikan keinoin on tutkittu.

## 3.2. Mitä metodilla on tutkittu

### 3.2.1. KIELIOPILLISTUMINEN

Kieliopillistuminen on ollut yksi keskeisimmistä historiallisen korpuslingvistiikan menetelmin tutkituista kielenmuutoksista. *Kieliopillistumisella* tarkoitetaan yleisesti ottaen ilmiötä, jossa olemassa olevan leksikaalisen sanan tai konstruktion merkitys muuttuu asteittain kieliopillisemmaksi: konstruktiot, joilla on aiemmin viitattu konkreettisiin asioihin tai tapahtumiin, valjastetaan kuvaamaan yhä abstraktimpia

merkityksiä, kuten aikamuotoja, aspektia tai lukua. Koska tällaiset kielioppikategoriat ovat pakollinen osa kielen järjestelmää siinä mielessä, että kielenkäyttäjän on tyypillisesti aina kytkettävä puhunnoksensa esimerkiksi oman kielensä tempus- tai aspektijärjestelmään valitsemalla tarkoitukseensa sopiva muoto, kieliopillistuneen konstruktion käyttötaajuus on valtavasti suurempi kuin mitä sillä oli ennen kieliopillistumista. Yleistyneen käytön seurauksena konstruktiossa tapahtuu usein fonologisia ja morfologisia muutoksia, joissa se muuttuu foneettisesti yksinkertaisemmaksi sekä mahdollisesti myös vapaasta morfeemista sidonnaiseksi morfeemiksi.

Esimerkiksi voidaan ottaa vaikkapa muinaisenglannin verbi *willan*. Se oli alkujaan leksikaalinen verbi, joka merkitsi 'tahtoa' tai 'haluta', kuten esimerkissä 1.

1.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *ða* | *cwæþ* | *he* | *to* | *him,* | ***wylt*** | *þu* | *hal* | *beon?* |
| sitten | sanoa.<br>3SG.PRET | hän | PREP | hän.<br>DAT.M.SG | haluta.<br>2SG.PRES | sinä | terve | olla/<br>tulla.<br>INF[22] |

'Sitten hän sanoi hänelle: haluatko tulla terveeksi?' (HC, 950–1050)

Vuosisatojen saatossa *willan*-verbiä alettiin käyttää yleisemmin ilmaisemaan tulevia, ennustettavia tapahtumia ja lopulta pelkästään tulevaa aikaa. Prosessin lopputuloksena verbistä tuli futuurisuutta ilmaiseva apuverbi (nykyenglannissa *will*), joka varsinkin puheessa esiintyy usein muodossa *'ll* (vrt. *we will*; *we'll*).

Toinen esimerkki kieliopillistumiselle ominaisista semanttisista ja rakenteellisista muutoksista ovat englannin kielen binominaaliset kvanttorit (esim. *a lot of* N, *a bit of* N). Sekä *lot* että *bit* ovat näissä rakenteissa viitanneet alkujaan konkreettisiin asioihin: *lot* maapalstaan (esim. *a lot of land*) ja *bit* haukkapalaan (esim. *a bit(e) of bread*). Kieliopillistumisprosessissa kielenpuhujat tulkitsivat näiden sanojen merkityksen uudelleen, niin että niiden tarkoitteeseen implisiittisesti liittynyt pieni/suuri

koko tai määrä nousikin keskeiseksi, ja tästä seurasi konstruktioiden rakenteellinen uudelleentulkinta: *lot*- ja *bit(e)*-sanat eivät enää olleetkaan substantiivilausekkeidensa pääsanoja, vaan ne käsitettiin syntaktisesti osaksi monisanaista kvanttorirakennetta. Tällöin alun perin jälkimääritteen osana olleesta substantiivista tuli uudelleen tulkitun lausekkeen pääsana: [a lot] [of land] → [a lot of] [land]. Kieliopillistumista voidaan tarkastella myös käyttökontekstien laajentumisen näkökulmasta. Esimerkiksi *a lot (of)*- ja *a bit (of)* -kvanttoreita on alettu käyttää myös muunlaisten kieliopillisten merkitysten ilmaisuun, kuten määrää kuvaavina adverbeina (esim. *I like you a lot*) ja asteadverbeina (esim. *a bit rude*). Samoin kuin *will*-apuverbistä myös näistä konstruktioista on olemassa kontrahoidut muodot: varsinkin *lotta* on yleinen, mutta myös *bitta*-muotoa käytetään etenkin puheessa. Sama koskee myös monia muita englannin binominaalisia kvanttoreita, kuten *sort of* (*sorta*), *kind of* (*kinda*) ja *type of* (*typa*).

Yksi tutkituimmista kieliopillistumisprosesseista on englannin *be going to* -konstruktion muuttuminen liikettä kuvaavasta rakenteesta futuuria ilmaisevaksi konstruktioksi. Esimerkit 2 ja 3 kuvaavat rakenteen käyttöä varhaisuusenglannissa juuri ennen kieliopillistumisprosessin alkua, kun taas 4 ja 5 ovat esimerkkejä nykyenglannissa yleisestä, kieliopillistuneesta futuurista. Esimerkki 5 on kiistaton todiste rakenteen kieliopillistumisesta: leksikaalinen *go*-verbi on lauseessa toistettu, koska *be going to* -rakenteessa olevaa *going*-sanaa ei enää käsitetä liikettä ilmaisevaksi verbiksi.

2. *A yong mayd* ***going to*** *a feast with hir mother*... (PPCEME[23], 1582)
3. *Mistris Page, trust me, I* ***was going to*** *your house.* (PPCEME, 1599)
4. *It***'s going to*** *be tough for them.* (COCA[24], 2012)
5. *I***'m going to*** *go to Nashville, Tennessee.* (COCA, 2012)

Esimerkit 2–5 ilmentävät *be going to* -konstruktion muutoksen alku- ja päätepisteitä, mutta rakenteen kieliopillistuminen tapahtui tosiasiassa vaiheittain vuosisatojen saatossa. Muutos sai alkunsa, kun kielenpuhujat alkoivat 1600-luvulla korostaa fyysiseen liikkeeseen tyypillisesti liittyvää tahdonalaisuutta ja aikomusta siinä määrin, että konstruktion ei

enää välttämättä tarvinnutkaan kuvata fyysistä liikettä. Esimerkeissä 6 ja 7 on toki mahdollista, että puhuja liikkuu paikasta toiseen puhumaan, mutta todennäköisempää on, että hän ilmaisee ainoastaan aikomuksensa puhua.

6. *now, a certaine religious Woeman dyed at home with vs, about a day and a halfe, before that occurred, whereof i* ***am going to*** *speake* (EEBO, 1642)
7. *but there is another thing worth your observation, which i* ***am going to*** *tell you* (EEBO, 1653)

Leuvenin yliopiston tutkijat Sara Budts ja Peter Petré tarkastelivat vuonna 2016 julkaistussa korpustutkimuksessaan *be going to* -konstruktion kieliopillistumista myöhäisuusenglannin kaudella (1700–1920). Myös he tulkitsivat *be going to* -rakenteen kieliopillistumisen subjektifikaatioprosessiksi, jossa konstruktion alkuperäinen, liikettä kuvaava objektiivinen merkitys muuttui ensin subjektiivisemmaksi (puhujan omaa intentiota tarkoittavaksi) ja myöhemmin vielä subjektiivisemmaksi, niin että puhujat alkoivat käyttää konstruktiota ennakoidessaan yleisemmin tulevia tapahtumia ja asioita tai ollessaan epävarmoja tulevista tapahtumista. Budts ja Petré havaitsivat muun muassa, että ensimmäisen persoonan subjektipronominien osuus kaikista *be going to* -rakenteiden subjekteista pieneni tutkitun kauden kuluessa, kun taas toisen persoonan subjektien osuus kasvoi. Tämä selittyy muun muassa kysymyslauseiden yleistymisellä: toiselle ihmiselle esitetyt kysymykset, kuten *are you going to come tomorrow?*, ovat esimerkki rakenteen uudenlaisesta käytöstä, jossa puhuja ei ilmaise omaa intentiotaan vaan osoittaa epävarmuutensa toisen henkilön aikomuksista. Budts ja Petré havaitsivat aineistossaan myös jyrkän laskun tulevan tapahtuman pikaisuudessa: vuosina 1710–1745 peräti 79 % *be going to* -rakenteista viittasi puhetilannetta välittömästi seuraavaan tapahtumaan, kun taas 1886–1920 vastaava osuus oli enää 29 %. Tilastollisena menetelmänä Budts ja Petré käyttivät Kendallin tau -testiä.

**Tietolaatikko 2.**

- Budtsin ja Petrén käyttämät korpukset olivat Penn Parsed Corpus of Modern British English (Kroch, Santorini & Diertani 2016), ECCO-TCP Corpus[25] ja Corpus of Late Modern English Texts v.3.0 (Diller, De Smet & Tyrkkö 2011).
- Useamman kuin yhden korpuksen käyttäminen historiallisessa korpustutkimuksessa on varsin yleistä; yksittäinen korpus voi olla liian suppea tarkasteltavan kielen ilmiön kattavaan tutkimiseen.
- Tutkijan on otettava tarkoin huomioon tutkimuksessaan käyttämiensä korpusten väliset erot. Koska suurin osa korpuksista on koottu eri periaatteita noudattaen, ei esimerkiksi sanojen frekvenssien vertailu korpusten välillä ole ongelmatonta.

Budtsin ja Petrén tutkimus on monessa mielessä tyyppiesimerkki historiallisen korpuslingvistiikan menetelmin toteutettavasta tutkimuksesta: tutkijat pureutuvat tarkastelemaansa ilmiöön analysoimalla kattavasti sen eri käyttökonteksteja ja niissä tapahtuvia mikrotason muutoksia, minkä ansiosta he pystyvät tarjoamaan yksityiskohtaisen analyysin kieliopillistumisprosessiin liittyvistä semanttisista tekijöistä.

### 3.2.2. Historiallinen sosiolingvistiikka: *-s* ja *-th* verbin kolmannen persoonan päätteinä englannin kielessä

Historiallinen sosiolingvistiikka on kielitieteen suuntaus, joka sai alkusysäyksensä sosiolingvististen hypoteesien testaamisesta historiallisissa aineistoissa. Näitä hypoteeseja olivat esimerkiksi naisten johtava rooli kielenmuutoksessa sekä ajatus siitä, että vertailemalla eri-ikäisten ihmisten kielenkäyttöä saadaan samalla tietoa kielenmuutoksesta ajassa: ihmisten kielenkäytön oletetaan suurelta osin vakiintuvan jo nuorena, joten vanhempien ihmisten kielenkäyttö edustaa varhaisempaa aikakautta kuin nuorempien (ns. näennäisaikahypoteesi). Suzanne

Romainen (1982) ohella historiallisen sosiolingvistiikan pioneereja ovat olleet Helsingin yliopiston tutkijat Terttu Nevalainen ja Helena Raumolin-Brunberg, joiden johdolla myös koottiin CEEC-korpus.

Nevalainen ja Raumolin-Brunberg (2003) tutkivat useisiin eri kielellisiin muuttujiin eli variaabeleihin pohjautuvia muutoksia CEEC-korpuksessa aikavälillä 1410–1681 historiallisen sosiolingvistiikan näkökulmasta. Otamme tähän esimerkiksi kuvassa 3 esitellyn englannin yksikön kolmannen persoonan preesensin verbipäätteen variaabelin -*s*/-*th*. Nevalainen ja Raumolin-Brunberg huomauttavat, että kyseessä on oikeastaan kaksi eri muutosta: toisaalta painottoman /e/-vokaalin kato päätteestä, jonka alkuperäinen muoto oli siis -*eth*, ja toisaalta viimeisen konsonantin muuttuminen sibilantiksi. Nevalainen ja Raumolin-Brunberg osoittavat, että molemmat muutokset alkoivat Pohjois-Englannista. Ensimmäisenä alkoi -*eth*:n muuttuminen -*es*:ksi (esim. *eateth* vs. *eates*, joskin vokaalia saatettiin kirjoittaa esimerkiksi *y*:llä), mikä näkyy kuvassa 3 pienenä nousuna 1400-luvun lopulla; tämä ei kuitenkaan kunnolla levinnyt etelämmäksi. Noin sata vuotta myöhemmin -*s*-variantti aloitti uuden nousun, joka pohjautui -*eth*-muodon ja vokaalittoman -*s*-muodon väliseen vaihteluun (esim. *eateth* vs. *eats*); tuon ajan kirjoittajat pitivätkin -*s*-muotoa -*eth*-muodon lyhenteenä. Tällä kertaa muutos eteni useimmissa murteissa loppuun asti, ja nykyenglannin puhujat tuntevat -*th*-variantin lähinnä Shakespearen näytelmistä ja Jaakko-kuninkaan aikaisesta raamatunkäännöksestä (1611), joka on edelleen käytössä joissakin kirkkokunnissa.

Nevalainen ja Raumolin-Brunberg (2003) tarkastelivat muutosta useiden sosiaalisten kategorioiden suhteen. Kirjoittajan asuinpaikan lisäksi variaabelin käyttöön – eli siihen, kumpaa varianttia kirjoittaja käytti enemmän – saattoivat vaikuttaa mm. hänen sukupuolensa, yhteiskunnallinen asemansa sekä suhteensa kirjeen vastaanottajaan. Menetelmänä Nevalainen ja Raumolin-Brunberg käyttivät variantin normalisoidun frekvenssin laskemista, kuten kuvassa 3, mutta he jakoivat aikakaudet sosiaalisten kategorioiden perusteella muodostettuihin alakorpuksiin. Sitten he vertailivat alakorpuksien normalisoituja frekvenssejä hypoteesitestauksen keinoin $\chi^2$-testillä. Yllä mainitun alueellisen vaihtelun lisäksi he havaitsivat, että muutosta johtivat -*s*:n uuden nousun osalta odotetusti naiset.

Nopeita *-s*:n omaksujia olivat etenkin säätyläisiä alhaisemmassa yhteiskunnallisessa asemassa olleet kirjoittajat sekä toisaalta sosiaaliset nousijat. Täydentääkseen tuloksiaan Nevalainen ja Raumolin-Brunberg toteuttivat monimuuttuja-analyysin VARBRUL-ohjelmalla, jonka eri versioita on käytetty yleisesti sosiolingvistisessä tutkimuksessa.[26] Monimuuttuja-analyysin avulla pystytään arvioimaan eri sosiaalisten kategorioiden suhteellista tärkeyttä muutoksessa. Tässä analyysissa kirjoittajan asuinalue ja sukupuoli nousivat tärkeiksi muutosta selittäviksi tekijöiksi, kun taas kirjoittajan suhteella kirjeen vastaanottajaan (perheenjäsen/ystävä vs. muu) ei havaittu olevan merkittävää vaikutusta.

## 4. Yhteenveto

Historiallinen korpuslingvistiikka tarjoaa tutkijalle monipuolisen menetelmän kielen varhaisempien vaiheiden ja kielenmuutoksen systemaattiseen tarkasteluun. Korpuslingvistin työssä korostuu tutkimuksen lähteinä käytettyjen korpusten tarkka tunteminen, mutta kuten olemme pyrkineet tuomaan tässä artikkelissa esille, aineiston tunteminen on vain ennakkoedellytys tutkimuksen tekemiselle. Kieli voi muuttua monesta syystä, ja historiallisen korpuslingvistin työkalupakkiin on hyvä kuulua esimerkiksi sosiolingvistiikan, pragmatiikan, kielikontaktien sekä kognitiivisen kielitieteen teorioiden hallinta. Yleisesti ottaen historialliselle korpuslingvistiikalle leimaa-antavaa on määrällisten ja laadullisten menetelmien yhdistäminen: tutkijan korpusaineistosta laskemien frekvenssien merkitsevyyttä voidaan esimerkiksi tarkastella tilastollisin menetelmin, ja niitä voidaan havainnollistaa monenlaisten visualisointien avulla, mutta tulosten tulkinta nojaa aina laadullisen tutkimuksen periaatteisiin – numerot itsessään eivät ole vielä tutkimustulos. Toisaalta on hyvä pitää mielessä, että korkealaatuista korpuslingvistiikkaa on mahdollista tehdä myös ilman sofistikoituneita tilastollisia menetelmiä: kielenmuutokset voivat olla niin selkeitä, että niitä on mahdollista seurata pelkkien frekvenssitietojen avulla.

Tässä artikkelissa annetut esimerkit historiallisen korpuslingvistiikan kentästä on tarkoitettu ensikatsaukseksi metodologian eri soveltamismahdollisuuksiin. On selvää, että menetelmää käytetään laajalti myös muunlaisten tutkimuskysymysten tarkasteluun: esimerkiksi historiallista semantiikkaa, pragmatiikkaa, fonologiaa ja kielen murrevaihtelua on kaikkia tutkittu monipuolisesti korpuslingvistiikan menetelmin. **Digitaalisten ihmistieteiden** yleistymisen myötä on todennäköistä, että niin korpuksia kuin muitakin digitaalisia kieliaineistoja on tulevaisuudessa tutkijan käytössä entistä enemmän ja nämä aineistot ovat yhä useammin saatavilla myös muussa kuin tekstimuodossa. Esimerkiksi vanhojen kirjeiden digitoidut kuvat voivat tarjota paljon sellaista tietoa tekstistä ja sen tuottamiseen liittyvistä olosuhteista, joita pelkkä tietokoneen ruudulla näkyvä tekstirivi ei tavoita. Kuten CEEC-korpuksen esimerkki osoittaa, myös tekstien kirjoittajiin liittyviä taustatietoja on mahdollista kytkeä korpusaineistoon, jolloin ne ovat hyödyllisiä paitsi kielentutkijoille myös historioitsijoille. Muitakin tällaisia aineistoja on jo olemassa (esim. London Lives[27]), joskin ne on tyypillisesti koottu pääasiassa historiantutkimuksen tarpeisiin, eikä niiden käyttö sellaisenaan ole täysin ongelmatonta korpuslingvistiikan näkökulmasta.

Tämän lyhyen katsauksen pyrkimyksenä on ollut tarjota joitakin näkökulmia historiallisen korpuslingvistiikan taustaan, perusolettamuksiin ja viimeaikaisiin kehityskulkuihin. Luvussa 3 esitetyt tapaustutkimukset ovat esimerkkejä yksittäisiin kysymyksiin pureutuvasta tutkimuksesta, jotka osana suurempaa kokonaisuutta voivat tarjota uutta tietoa esimerkiksi sellaisista perustavanlaatuisista kysymyksistä, joita nostimme esiin artikkelimme johdannossa. Toivomme, että tämä artikkeli välittää lukijalle kuvan korpuslingvistiikasta monipuolisena menetelmänä, joka kehittyy niin kielitieteen tutkimusalan kuin yleisen tietoteknisen kehityksen rinnalla ja jonka avulla voimme saada jatkuvasti uutta tietoa paitsi kielestä järjestelmänä myös ihmisyhteisöjen kielikäytäntöjen vaihtelusta ja muutoksesta.

**Aiheesta lisää:**

Hopper, Paul J. & Traugott, Elizabeth Closs. 2003. *Grammaticalization*. (2. p.). Cambridge: Cambridge University Press.

Krug, Manfred & Schlüter, Julia (toim.). 2013. *Research Methods in Language Variation and Change*. Cambridge: Cambridge University Press.

McEnery, Tony & Hardie, Andrew. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Nevalainen, Terttu & Raumolin-Brunberg, Helena. 2003. *Historical Sociolinguistics: Language Change in Tudor and Stuart England*. Longman Linguistics Library. London: Pearson Education.

## VIITTEET

1 http://www.helsinki.fi/varieng/CoRD/corpora/BROWN/.
2 http://www.helsinki.fi/varieng/CoRD/corpora/FROWN/.
3 http://www.helsinki.fi/varieng/CoRD/corpora/LOB/.
4 http://www.helsinki.fi/varieng/CoRD/corpora/FLOB/.
5 http://www.helsinki.fi/varieng/CoRD/corpora/BLOB-1931/.
6 http://www.helsinki.fi/varieng/CoRD/corpora/B-BROWN/.
7 http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/.
8 https://www.english-corpora.org/coha/.
9 https://www.english-corpora.org/eebo/, https://digi.kansalliskirjasto.fi.
10 http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/.
11 http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/, http://kaino.kotus.fi/korpus/vks/meta/vks_coll_rdf.xml.
12 https://www.kielipankki.fi.
13 http://www.meta-share.org, https://www.clarin.eu/content/language-resource-inventory, http://www.helsinki.fi/varieng/CoRD/.
14 https://www.corpusdelespanol.org, https://www.corpusdoportugues.org.
15 https://korp.csc.fi.
16 https://cqpweb.lancs.ac.uk/.
17 http://www.lexically.net/wordsmith/, http://www.laurenceanthony.net/software/antconc/.
18 https://www.r-project.org.
19 https://www.ibm.com/analytics/spss-statistics-software.

20 http://www.helsinki.fi/varieng/CoRD/corpora/CEEC/.
21 https://ggplot2.tidyverse.org.
22 Lyhenteet: 3SG = yksikön 3. persoona, PRET = preteriti, PREP = prepositio, DAT = datiivi, M = maskuliini, 2SG = yksikön 2. persoona, PRES = preesens, INF = infinitiivi.
23 PPCEME = Penn-Helsinki Parsed Corpus of Early Modern English (https://www.ling.upenn.edu/hist-corpora/PPCEME-RELEASE-3/).
24 COCA = Corpus of Contemporary American English (https://www.english-corpora.org/coca/).
25 https://textcreationpartnership.org/tcp-texts/ecco-tcp-eighteenth-century-collections-on-line/.
26 https://en.wikipedia.org/wiki/Variable_rules_analysis.
27 https://www.londonlives.org.

## KIRJALLISUUS


Baker, Helen, Brezina, Vaclav & McEnery, Tony. 2017. Ireland in British Parliamentary debates 1803–2005: Plotting changes in discourse in a large volume of time-series corpus data. Julkaisussa: Säily, Tanja, Nurmi, Arja, Palander-Collin, Minna & Auer, Anita (toim.) *Exploring Future Paths for Historical Sociolinguistics*. Advances in Historical Sociolinguistics 7. Amsterdam: John Benjamins, 83–107.

Baron, Alistair. 2011. VARD 2. Tietokoneohjelma. Saatavissa: http://ucrel.lancs.ac.uk/vard/.

Benveniste, Emile. 1971 [1958]. Subjectivity in language. Julkaisussa: *Problems in General Linguistics*. Englanninkielinen käännös Mary Elizabeth Meek. Coral Gables (FL): University of Miami Press, 223–230.

Biber, Douglas. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, Douglas & Finegan, Edward. 1997. Diachronic relations among speech-based and written registers in English. Julkaisussa: Nevalainen, Terttu & Kahlas-Tarkka, Leena (toim.) *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen*. Mémoires de la Société Néophilologique de Helsinki 52. Helsinki: Société Néophilologique, 253–275.

Budts, Sara & Petré, Peter. 2016. Reading the intentions of *be going to*: On the subjectification of future markers. *Folia Linguistica Historica* 37:1, 1–31.

Coolidge, Frederick L. 2013. *Statistics: A Gentle Introduction*. (3. p.). Thousand Oaks (CA): Sage.

Diller, Hans-Jürgen, De Smet, Hendrik & Tyrkkö, Jukka. 2011. A European database of descriptors of English electronic texts. *The European English Messenger* 19, 21–35.

Francis, W. Nelson & Kučera, Henry. 1979. *Manual to Accompany a Standard Sample of Present-day Edited American English, for Use with Digital Computers*. Alkuperäinen p. 1964, korjattu 1971, korjattu ja täydennetty 1979. Providence, Rhode Island: Department of Linguistics, Brown University.

Gries, Stefan Th. 2013. *Statistics for Linguistics with R: A Practical Introduction*. (2. p.). Berlin: De Gruyter Mouton.

Gries, Stefan Th. & Hilpert, Martin. 2012. Variability-based Neighbor Clustering: A bottom-up approach to periodization in historical linguistics. Julkaisussa: Nevalainen, Terttu & Traugott, Elizabeth Closs (toim.) *The Oxford Handbook of the History of English*. Oxford Handbooks in Linguistics. Oxford: Oxford University Press, 134–144.

Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä, Ulla & Lounela, Mikko (toim.). 2012. *Genreanalyysi – tekstilajitutkimuksen käytäntöä*. Kotimaisten kielten keskuksen verkkojulkaisuja 29. [verkkoaineisto]. [viitattu 7.2.2018]. Helsinki: Kotimaisten kielten keskus. Saatavissa: http://kaino.kotus.fi/www/verkkojulkaisut/julk29/.

Hilpert, Martin. 2011. Dynamic visualizations of language change: Motion charts on the basis of bivariate and multivariate data from diachronic corpora. *International Journal of Corpus Linguistics* 16:4, 435–461.

Kampstra, Peter. 2008. Beanplot: A boxplot alternative for visual comparison of distributions. *Journal of Statistical Software* 28, Code Snippet 1. Saatavissa: http://www.jstatsoft.org/v28/c01/.

Kroch, Anthony, Santorini, Beatrice & Diertani, Ariel. 2016. *The Penn Parsed Corpus of Modern British English* (PPCMBE2). Department of Linguistics, University of Pennsylvania. CD-ROM, 2. p., versio 1. Saatavissa: http://www.ling.upenn.edu/ppche/ppche-release-2016/PPCMBE2-RELEASE-1.

Lijffijt, Jefrey, Säily, Tanja & Nevalainen, Terttu. 2012. CEECing the baseline: Lexical stability and significant change in a historical corpus. Julkaisussa: Tyrkkö, Jukka, Kilpiö, Matti, Nevalainen, Terttu & Rissanen, Matti (toim.) *Outposts of Historical Corpus Linguistics: From the Helsinki Corpus to a Proliferation of Resources*. Studies in Variation, Contacts and Change in English 10. Helsinki: VARIENG. [verkkoaineisto]. [viitattu 7.2.2018]. Saatavissa: http://www.helsinki.fi/varieng/series/volumes/10/lijffijt_saily_nevalainen/.

Marttila, Ville. 2014. *Creating Digital Editions for Corpus Linguistics: The Case of Potage Dyvers, a Family of Six Middle English Recipe Collections*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-0060-3.

McEnery, Tony & Hardie, Andrew. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Metsämuuronen, Jari. 2011. *Tutkimuksen tekemisen perusteet ihmistieteissä: tutkijalaitos*. (E-kirjan 1. p.). Helsinki: International Methelp.

Nevalainen, Terttu & Raumolin-Brunberg, Helena. 2003. *Historical Sociolinguistics: Language Change in Tudor and Stuart England*. Longman Linguistics Library. London: Pearson Education.

Romaine, Suzanne. 1982. *Socio-historical Linguistics: Its Status and Methodology*. Cambridge Studies in Linguistics. Cambridge: Cambridge University Press.

Szmrecsanyi, Benedikt, Rosenbach, Anette, Bresnan, Joan & Wolk, Christoph. 2014. Culturally conditioned language change? A multi-variate analysis of genitive constructions in ARCHER. Julkaisussa: Hundt, Marianne (toim.) *Late Modern English Syntax*. Cambridge: Cambridge University Press, 133–152.

Säily, Tanja. 2014. *Sociolinguistic Variation in English Derivational Productivity: Studies and Methods in Diachronic Corpus Linguistics*. Mémoires de la Société Néophilologique de Helsinki XCIV. Helsinki: Société Néophilologique. Saatavissa: http://urn.fi/URN:ISBN:978-951-9040-50-9.

Vartiainen, Turo, Säily, Tanja & Hakala, Mikko. 2013. Variation in pronoun frequencies in early English letters: Gender-based or relationship-based? Julkaisussa: Tyrkkö, Jukka, Timofeeva, Olga & Salenius, Maria (toim.) *Ex Philologia Lux: Essays in Honour of Leena Kahlas-Tarkka*. Mémoires de la Société Néophilologique de Helsinki XC. Helsinki: Société Néophilologique, 233–255.
Ware, Colin. 2004. *Information Visualization: Perception for Design*. (2. p.). San Francisco: Morgan Kaufmann.

# Osa III

# Kielellisen monimuotoisuuden selvittelyyn soveltuvia menetelmiä

# Luku 12 Kielikontaktien tutkimus

*Max Wahlström*
https://orcid.org/0000-0002-9585-5217
*Jouko Lindstedt*
https://orcid.org/0000-0001-9021-7622

**Mikä?**

Kielikontaktien tutkimus tarkastelee kahden tai useamman kielimuodon kohtaamiseen liittyviä ilmiöitä. Usein tutkimuskohteena on kielikontaktin aikaansaama kielenmuutos, mutta huomio voi olla myös esimerkiksi monikielisiin tilanteisiin liittyvissä ilmiöissä ja käytänteissä. Kielikontakti voi vaikuttaa kaikkiin kielen tasoihin, ja siksi tutkimuksessa käytetyt menetelmät ovat usein samoja kuin muilla kielitieteen aloilla. Kontakti-ilmiöiden ymmärtäminen ja kontaktilähtöisen kielenmuutoksen osoittaminen vaativat tyypillisesti useiden eri menetelmien soveltamista.

**Katso myös:**

Luku 14 Kielitypologian menetelmät

Luku 15 Historiallinen ja vertaileva kielentutkimus

Luku 16 Antropologinen lingvistiikka

Luku 21 Sosiolingvistiikka

# 1. Johdanto

Kielikontakti tarkoittaa laajimmillaan mitä tahansa tilannetta, jossa on tavalla tai toisella läsnä enemmän kuin yksi kielimuoto, olivatpa sitten kyseessä puhutut, viitotut tai kirjoitetut kielet. Useimmiten kielikontaktien tutkimuksen kiinnostuksen kohteena ovat ainakin jossain määrin toistuvat monikieliset tilanteet ja tällaisten tilanteiden kielelliset seuraukset eli kontaktilähtöinen kielen muutos.

Kielikontaktin laajimman määritelmän kannalta on vaikea sanoa, mihin kielentutkimuksen alaan kielikontaktit eivät millään tavalla liittyisi. Vaikka ajateltaisiin yksikielistä puhujaa, joka on tekemisissä vain samaa kieltä puhuvien ihmisten kanssa, hän altistuu mahdollisesti myös toisille murteille. Hän käyttää todennäköisesti myös lainasanoja, vaikka kyseinen yhteisö ei kyseisellä hetkellä olisikaan tekemisissä muiden kielten puhujien kanssa. Lisäksi puhujan kieli muistuttaa todennäköisesti rakenteeltaan muita samalla alueella puhuttuja kieliä, vaikka ne eivät olisikaan sille sukua. Lainasanasto ja tietyille maantieteellisille alueille tyypilliset, kielikuntarajat ylittävät rakennepiirteet kertovat historiallisista kontakteista. Turvallisinta onkin olettaa, että jokainen kieli on historiansa aikana ollut alttiina kielikontakteille ja että kielen kulloinenkin asu on aina jossain määrin kontaktien muovaama.

Kielikontaktien ja niiden seurausten esiintyminen lähes kaikkialla ei merkitse kuitenkaan sitä, että niiden rooli olisi helppo osoittaa esimerkiksi kielen muutosta selitettäessä. Kontaktilähtöisiä selityksiä on lisäksi vieroksuttu aiemmassa historiallisessa kielentutkimuksessa: kontaktiselitys on kelvannut usein vain siinä tapauksessa, että muutosta ei ole voitu selittää kielensisäisillä tekijöillä. Synkroniselle kielentutkimukselle voi lisäksi olla melko yhdentekevää jonkin piirteen tai sanan alkuperä, eritoten mikäli ilmiön kontaktilähtöisyys ei ole puhujien yleisessä tiedossa. Tosin mikäli jonkin piirteen alkuperä kuitenkin on tiedossa esimerkiksi siitä syystä, että puhujat tunnistavat sen selkeästi lainatuksi jostakin tietystä kielestä, arkiajattelussa tällaiseen piirteeseen voi liittyä esimerkiksi kielteisiä asenteita. Myös kielen normittamiseen on lähes

aina liittynyt puristisia pyrkimyksiä, joiden tarkoitus on nimenomaisesti puhdistaa vieraaksi katsottu aines.

Kielikontakteja tutkitaan monella kielitieteen alalla. Karkeasti lähestymistapoja voi rajata sen mukaan, keskittyykö tutkimus synkronisesti havainnoitaviin ilmiöihin vai aiemmin tapahtuneiden kontaktien jättämiin jälkiin kielissä. Lisäksi kielikontaktien roolin arviointi on useimmiten olennainen osa niin historiallista kielentutkimusta kuin kielitypologiaa (ks. Historiallinen ja vertaileva kielentutkimus; Kielitypologian menetelmät tässä kirjassa). Lähestymistapojen moninaisuudesta johtuen kielikontaktien tutkimuksessa käytettyjä metodeja onkin lukuisia, ja vain osa niistä on käytössä yksinomaan kontaktien tutkimuksen piirissä. Tämä käy ilmi esimerkiksi merkittävimmän laaja-alaisen kielikontaktien tutkimuksen aikakausjulkaisun *Journal of Language Contact* artikkelikirjosta.

Seuraavassa luvussa esitellään kielikontaktien tutkimuksen historiaa. Kielikontaktitutkimuksen monialaisuuden vuoksi menetelmiä ja niiden valintaa lähestytään tutkittavien ilmiöiden ja alan teoreettisten peruskysymysten kautta. Kolmannessa luvussa käsitellään kielikontakteja kielenmuutoksen selittäjänä (3.1.), kontaktitilanteiden tutkimusta (3.2.), kontaktin historian arviointia sen tuloksen kautta (3.3.), areaalilingvistiikkaa eli kontaktialueiden tutkimusta (3.4.), kontaktin synnyttämiä uusia kielimuotoja (3.5.) ja kontaktien tutkimuksen sekä typologian suhdetta (3.6.). Lopuksi esitetään kootusti, mitä menetelmällisiä valmiuksia kielikontaktien tutkijalta vaaditaan.

## 2. Kielikontaktien tutkimuksen historia ja sen asema historiallisessa kielentutkimuksessa

Kielikontaktien tutkimuksen historia ulottuu paljon kauemmaksi menneisyyteen kuin itse käsite. Esimerkiksi sanaston lainautuminen tunnettiin ilmiönä jo antiikissa, vaikka etymologinen tutkimus ymmärrettiin pitkään varsin toisella tapaa kuin nykyään: etymologiassa oli kyse sanojen merkityksen ja alkuperän selittämisestä usein varsin

vaikutelmanomaisesti, ja vasta 1800-luvulla lainasanoista tuli systemaattisen tutkimuksen kohde. Lainasanatutkimusta voidaankin pitää ensimmäiseksi vakiintuneena kielikontaktien tutkimuksen suuntauksena.

Lainasanatutkimuksen kehitys on tiiviisti yhteydessä nuorgrammaattisen historiallis-vertailevan kielitieteen syntyyn 1800-luvulla. Säännöllisten äännevastaavuuksien löytyminen sukukielten välillä mahdollisti niin kielten sukulaissuhteiden toteamisen kuin kantakielten rekonstruktion. Samankaltainen säännöllinen vastaavuus äänteiden välillä havaittiin myös lainasanastossa: tietystä kielestä tiettyyn aikaan lainautuneiden sanojen äänteistö korvataan lainansaajakielessä aina tietyillä äänteillä. Toisaalta lainasanat on pystyttävä erottamaan niistä sanoista, jotka osoittavat kielisukulaisuutta. Lainasanatutkimus ja historiallinen kielentutkimus muotoutuivat pian symbioottisiksi: niin äänteistön rekonstruktion kuin kielen muutoksen suhteellisen ja absoluuttisen ajoituksen osalta lainasanojen tuntemuksella on keskeinen merkitys (ks. Historiallinen ja vertaileva kielentutkimus tk.).

Historiallis-vertailevan kielentutkimuksen saavutukset, kuten uusien kielikuntien löytyminen ja kielten välisten suhteiden mallintaminen sukupuuna, toivat mukanaan myös havaintoja, joiden selittämiseen äännevastaavuuksiin perustuva metodiikka ei pystynyt. Eräs näistä havainnoista liittyi kolonialismin synnyttämiin kielitilanteisiin. Joidenkin, usein afrikkalaisten tai afrikkalaistaustaisten yhteisöjen kieleksi oli muodostunut sanastoltaan eurooppalaisten kolonisoijien kieliin pohjautuva mutta kieliopiltaan uudelleenjärjestynyt kreoli, joka oli usein syntynyt usean kielen kontaktitilanteessa. 1800-luvun loppupuolelle asti kielitieteilijöiden suhtautuminen kreoleihin heijasteli tuon ajan rasistista ja etnosentristä ihmiskäsitystä: kreolit erosivat leksifioijakielistään esimerkiksi vähäisemmällä sanataivutuksellaan, mikä tulkittiin muun muassa kreolien puhujien älylliseksi kykenemättömyydeksi omaksua kehittyneempinä pidettyjen eurooppalaiskielten kielioppia (ks. myös luku 3.6.).

Kontaktikieliksi kutsuttujen kielimuotojen, kuten juuri kreolien mutta myös pidginien ja sekakielten (sa. *Mischsprache,* engl. *mixed language*), tutkimuksen uranuurtajana voidaan pitää saksalaista Hugo Schuchardtia (1842–1927; ks. Meijer & Muysken 1977; kontaktikielistä ks. luku 3.5.). Schuchardt tunnetaan myös nuorgrammaatikkojen

sukupuumallille vaihtoehtoisen aaltomallin kehittäjänä. Vaikka aaltomalli kunnianhimoisimmassa muodossaan unohtuikin, Schuchardtin nuorgrammaatikkoihin kohdistamaan kritiikkiin liittyi elementtejä, jotka 1900-luvun lopulla saivat laajan hyväksynnän. Parhaiten näistä tunnetaan Schuchardtin ajalleen vastavirtainen näkemys, että kaikki kielet ovat enemmän tai vähemmän ”sekakieliä”, toisin sanoen on oletettava, että jokainen kieli on ollut jonakin aikana alttiina kielikontaktille (ks. esim. Nicolaï 2014, 228–232).

Vuosikymmentä myöhemmin puolalaissyntyinen Jan Baudouin de Courtenay (1845–1929) pohti, Schuchardtin hengessä, kaikkien kielten kontaktiluonnetta (Boduen-de-Kurtenè 1901). Hänen tarkastelukulmansa ei liittynyt kuitenkaan historialtaan läpinäkyviin kontaktikieliin, kuten kreoleihin, vaan ilmiöihin, joita nykyään kutsutaan areaalilingvistisiksi. Kielitieteellisen kiinnostuksen kohdistuessa yhä useampaan kieleen ja kielikuntaan havaittiin, että joillakin alueilla puhutut kielet muistuttivat toisiaan merkittävästi, vaikka ne eivät olleet lähisukukieliä tai sukua ollenkaan. De Courtenayn yhtenä kiinnostuksen kohteena olleiden Balkanin kielten (kreikan, albanian, eteläslaavilaisten ja Balkanin romaanisten kielten) osalta tällaisia havaintoja esitettiin jo 1800-luvun alussa (Kopitar 1829). Pohjois-Amerikassa vastaavia havaintoja kielipiirteiden leviämisestä teki Franz Boas (1858–1942; ks. esim. 1911, 47–53) mantereen luoteisosan alkuperäiskielissä.

Tällaisten toisiaan muistuttavien naapurikielten ryhmien nimitykseksi vakiintui *kieliliitto* (sa. *Sprachbund* < ven. *jazykovoj sojuz*) Nikolai Trubetzkoyn (1890–1938; Trubetzkoy 1928, 18) ensimmäisessä kansainvälisessä lingvistikongressissa esittelemän ehdotuksen myötä. Trubetzkoyn ehdotuksessa *kieliliitto* ja *kielikunta* (sa. *Sprachfamilie*) olivat *kieliryhmän* (sa. *Sprachgruppe*) alakäsitteitä. Siinä missä kielikunta määrittyy kielisukulaisuuden kautta, tällaista yhteyttä ei kieliliiton kielillä ole. Määritelmässä kieliliiton kielet muistuttavat toisiaan suuressa määrin syntaksiltaan ja morfologiselta rakenteeltaan. Niissä on merkittävä määrä yhteisiä kulttuurisanoja ja mahdollisesti myös pinnallisia äänteellisiä samankaltaisuuksia, erona kielikunnissa esiintyviin yhteiseen perussanastoon ja säännöllisiin äännevastaavuuksiin. Usein on esitetty näkemyksiä, joiden mukaan Trubetzkoyn määritelmän taustalla olisi

ollut lähinnä Balkan parhaiten tuolloin tunnettuna kieliliittona. Kuitenkin esitellessään käsitteen ensimmäistä kertaa Trubetzkoy (1923) mainitsee Balkanin ainoastaan alaviitteessä keskittyen hahmottelemiinsa uralilais-altailaiseen ja Välimeren kieliliittoihin, joita nykyään pidettäisiin kokonsa vuoksi pikemmin makroalueina.

**Eurooppa kontaktialueena**

Euroopassa puhuttavat indoeurooppalaiset kielet olivat pitkään se normi, johon muun maailman "eksoottisia" kieliä eksplisiittisesti tai implisiittisesti verrattiin. Tilanne oli nurinkurinen, sillä Euroopan kielet edustavat vain noin 1–2 prosenttia maailman kielistä – kahden prosentin suuruusluokkaan pääsemiseksi on otettava mukaan myös Kaukasian kielet. Kielitypologisen tutkimuksen myötä on noussut esiin se käsitys, että Euroopan kielet ovatkin varsin epätavallisia koko maailman mittakaavassa ja että Eurooppa, varsinkin Länsi- ja Keski-Eurooppa, on tiivis kielellinen alue, jossa kielet ovat lainanneet toisiltaan piirteitä kontaktien kautta. Jotkut tutkijat puhuvat jopa Euroopan kieliliitosta samaan tapaan kuin Balkanin kieliliitosta.

Euroopan Tiedesäätiö rahoitti vuosina 1990–1994 laajaa EUROTYP-projektia. Sen tulokset julkaistiin kahdeksassa paksussa niteessä, joissa Euroopan kieliä on kuvattu typologisesti monesta eri näkökulmasta. EUROTYP-projektiin keskeisesti osallistunut Martin Haspelmath esitti teesin Euroopasta kielellisenä alueena tai jopa kieliliittona, jolle hän antoi nimen SAE eli *Standard Average European* (Haspelmath 1998; 2001; SAE oli alun perin Benjamin Lee Whorfin 1930-luvulla kehittämä termi, jolla hän kuvasi eurooppalaisten kielten välittämän käsiterakenteen samankaltaisuutta, ks. esim. Whorf 2012, 178). Euroopan kielten yhteisiä piirteitä ovat käsitelleet myös Heine ja Kuteva (2006).

Haspelmathin esittämiä piirteitä ovat esimerkiksi määräinen ja epämääräinen artikkeli, relatiivipronominilla muodostettava relatiivilause, 'omistaa'-apuverbiä käyttävä perfektitempus ja partisiippien avulla muodostettava passiivilause. Millä metodeilla tällaisten piirteiden leviämistä

kielikontaktien kautta voidaan tarkastella ja perustella? Tässä voidaan erottaa kolme vaihetta:

1. Ensin on osoitettava, että näitten piirteitten esiintyminen alueen kielissä on suurempi kuin pelkän sattuman kautta voitaisiin odottaa. Tähän on nykyään hyvänä apuna verkossa konsultoitavissa oleva *The World Atlas of Language Structures* (WALS): vaikka siitä aina väliin paljastuu virheitä yksittäisten kielten kuvauksissa ja isompien kielten osalta kuvaukset perustuvat kirjakieliin, se antaa kuitenkin hyvän kuvan siitä, mikä on yleistä, mikä harvinaista ja millainen maantieteellinen jakauma eri piirteillä on. Euroopan kielistä löytyy tosiaan monia yleisiä rakennepiirteitä, jotka erottavat ne kartalla selvästi naapurialueista (Keski-Aasiasta, Lähi-idästä ja Pohjois-Afrikasta) tai jopa koko maailmasta, joten ei ole järkevää olettaa, että jokainen Euroopan kieli olisi kehittänyt ne toisista riippumatta.
2. Sen jälkeen on osoitettava, etteivät yhteiset piirteet ole kielisukulaisuuden aiheuttamia. Tämähän voisi hyvinkin olla mahdollista, koska suurin osa Euroopan kielistä kuuluu indoeurooppalaiseen kielikuntaan ja Haspelmathin piirteet ovat keskittyneet niihin, vieläpä nimenomaan romaanisiin ja germaanisiin kieliin; slaavilaisissa kielissä ja esimerkiksi uralilaisissa kielissä unkarissa ja suomessa piirteitä on vähemmän. Kielihistoriallinen tieto osoittaa kuitenkin, ettei näitä piirteitä ollut indoeurooppalaisessa kantakielessä (tai missään Euroopan indoeurooppalaisten kielten myöhemmässä yhteisessä kantakielessä) ja että useat niistä kehittyivät ja levisivät vasta ajanlaskumme ensimmäisellä vuosituhannella, kun indoeurooppalaiset kielet olivat jo lähteneet kukin omalle kehitystielleen. SAE-piirteiden keskittyminen Eurooppaan ei siis johdu kielisukulaisuudesta. (Sukukieltenkin yhteinen piirre voi siis olla kontaktilähtöinen, jos sitä ei esiintynyt niiden yhteisessä kantakielessä.)
3. Vaiheet 1–2 riittävät osoittamaan piirteet kontaktilähtöisiksi, mutta hypoteesi saa vielä lisävahvistusta, jos piirteitten lainautumisen mekanismia ja sen sosiaalista kontekstia voidaan kuvata tarkemmin. Haspelmath esittää useitten (vaikkei kaikkien) SAE-piirteiden levinneen kielestä toiseen vanhan ajan ja keskiajan taitteessa kansain-

vaellusaikana, jolloin Euroopan väestölliset ja valtiolliset olosuhteet muuttuivat radikaalisti. Euroopan kontaktialueen ydintä eli SAE-piirteiden keskittymää Länsi- ja Keski-Euroopassa kutsutaankin joskus ”Kaarle Suuren kieliliitoksi” (engl. *Charlemagne sprachbund*, van der Auwera 1998, 824).

Samanaikaisesti kun havaintoja kontakti-ilmiöistä alettiin esitellä, nousi esiin kysymys kontaktilähtöisen kielenmuutoksen mekanismeista. Italialainen Graziadio Isaia Ascoli (1829–1907) kehitteli 1800-luvun loppupuoliskolla niin kutsutun substraattiteorian, jonka mukaan murteiden eriytymisen taustalla on usein kielenvaihto (ks. esim. Pellegrini 1980, 45). *Substraatti* on osa Ascolin luomaa käsitteistöä kielikontaktin sosiolingvistisille asetelmille: Substraatti on syrjäytynyt kieli, jonka vaikutus on nähtävissä sen syrjäyttäneessä kielessä, ja tyypillisesti substraatti havaitaan äänteistössä ja paikannimistössä. *Adstraatilla* puolestaan tarkoitetaan kieltä, joka on vaikuttanut toiseen kieleen kummankaan syrjäytymättä. Usein oletetaan myös, että adstraattisuhteessa olevat kielet ovat prestiisiltään eli arvostukseltaan samankaltaista. *Superstraatti* on sosiaalisesti hallitseva kieli, joka vaikuttaa sosiaalisesti alempaan kieleen; jotkut käyttävät termiä tilanteessa, jossa superstraatti syrjäyttää substraatin, jotkut taas tilanteessa, jossa superstraatti itse lopulta katoaa toisen kielen tieltä sosiaalisesti korkeammasta asemastaan huolimatta. Superstraattina voi toimia esimerkiksi *lingua franca*, eri kielten puhujien keskinäisessä kommunikaatiossa käyttämä yleiskieli kuten englanti nykyään suuressa osassa maailmaa. Toinen italialainen kielitieteilijä, Matteo Bártoli (1873–1946), puolestaan kehitti malleja kielellisten innovaatioiden leviämisestä maantieteellisesti ja prestiisin merkityksestä näiden omaksumisessa (Hall 1946).

Kielikontaktien mekanismien tutkimus monipuolistui 1900-luvun loppua kohti. Italialaisten kollegoidensa tavoin nykyaikaisen kielitieteen yksi uranuurtaja, Leonard Bloomfield (1887–1949), pyrki tarkastelemaan kontaktilähtöisen kielenmuutoksen vuorovaikutuksellisia ulottuvuuksia niin eri kielten kuin kielen varieteettien välillä (ks. esim. Bloomfield

1933, 445–495). Sosiolingvistiikan nousu toi mukanaan tarkempia malleja kieliyhteisöiden tarkasteluun, ja kielikontaktitutkimuksen klassikossa, Uriel Weinreichin (1926–1967) teoksessa *Languages in Contact* (1970 [1953]), keskeisellä sijalla ovat havainnot toisen kielen oppimisesta ja monikielisyydestä. 1900-luvun lopulle tultaessa asemansa olivat vakiinnuttaneet myös sellaiset lähtökohdiltaan synkronista kielikontaktien tutkimusta edustavat alat kuin koodinvaihdon tutkimus ja sosiolingvistinen monikielisyyden tutkimus (ks. luku 3.2.).

# 3. Miten kielikontakteja tutkitaan

Kielikontaktien tutkimuksessa käytetyt menetelmät riippuvat siitä, millaisia kontakti-ilmiöitä havainnoidaan ja mitä kontaktilähtöisen muutoksen osa-aluetta on tarpeen kuvata ja selittää. Tämä luku esittelee kontaktitutkimuksen metodologiaa ilmiöiden ja niiden selittämisen lähtökohdista.

## 3.1. Millaiset muutokset voivat olla kontaktilähtöisiä

Kielitieteessä oli pitkään vallalla käsitys, jonka mukaan kieli voi omaksua toisesta vain sellaisia piirteitä, jotka jollakin tapaa sopivat sen rakenteeseen ja omiin kehitystendensseihin; tällaisia käsityksiä ovat esittäneet esimerkiksi Antoine Meillet, Edward Sapir ja Roman Jakobson (Thomason 2001, 63). Näyttää kuitenkin siltä, että mille tahansa tällaiselle *strukturaaliselle* rajoitukselle löytyy vastaesimerkkejä, jos kontaktitilanne on *sosiolingvistisesti* tarpeeksi kontaktia suosiva. Tämän osoittivat erityisesti Sarah Thomason ja Terrence Kaufman (1988, 13–34) paljon siteeratussa kielikontakteja käsittelevässä monografiassaan (vrt. myös Thomason 2001, 63–66).

Tästä huolimatta on totta, että rakenteeltaan samankaltaiset kielet lainaavat toisiltaan piirteitä kielen eri tasoilla helpommin kuin

typologisesti kovin erilaiset. Sen lisäksi kielen tietyillä osa-alueilla kontaktivaikutus on todennäköisempää kuin toisilla. Kontaktivaikutukselle herkintä on sanasto mutta myös fonologia. Niinpä esimerkiksi suomi on lainannut runsaasti sanoja ruotsista. Suomenruotsi on puolestaan lainannut suomesta enemmän sanoja kuin ruotsinruotsalaiset murteet, ja se on myös lähestynyt suomea ääntämiseltään – siinä ei esimerkiksi esiinny juurikaan tooniaksenttia eikä klusiilien aspiraatiota (Kuronen & Leinonen 2001). Seuraavaksi herkimmin kieli ottaa vaikutteita toisista kielistä syntaksissa. Esimerkki aivan tuoreesta syntaktisesta lainavaikutuksesta englannista suomeen on konjunktion *koska* käyttö ilman predikaattiverbiä, kausaalisen preposition tapaan, esim. *en tehnyt läksyjä koska harkat.* Vastaava englannin *because*-sanan käyttö valittiin American Dialect Societyn "vuoden sanaksi" 2013 (Steinmetz 2014), joten piirre on lainattu suomeen aivan tuoreeltaan.

Vaikeammin tutkittavia mutta ilmeisesti varsin helposti tapahtuvia ovat kontaktilähtöiset muutokset semantiikassa, varsinkin yksittäisten sanojen merkityksessä, erityisesti jos puhujat pitävät kahden eri kielen sanoja samoina. Niinpä suomen *kontrolli* on perinteisen merkityksensä 'valvonta, tarkistus' lisäksi saanut englannin vaikutuksesta myös merkityksen 'hallinta, ohjaus' (*menetin pyörän kontrollin alamäessä*), ja *tieteen ja tekniikan kehityksen* asemesta puhumme yhä useammin *tieteen ja teknologian kehityksestä,* koska englannin *technique* 'taituruus; menetelmä' on merkitykseltään suppeampi kuin suomen *tekniikka*.

Vaikeimmin lainautuu sidonnainen morfologia eli erilaiset affiksit. Johtimet lainautuvat kuitenkin helpommin kuin kieliopilliset affiksit; esimerkiksi ruotsista suomeen lainatun -*(U)ri*-johtimen (*nahkuri, kankuri, suursyömäri*) alkuperä on viime kädessä latinassa (-*ārius*), josta se on levinnyt moniin eurooppalaisiin kieliin. Taivutusmorfologiassa on tavallisempaa, että jokin morfologinen kategoria lainautuu kuin että myös sen merkitsin lainautuu. Balkanin kieliliiton kielissä (ks. luku 3.4.) on paljon samankaltaisia morfosyntaktisia kategorioita mutta lainautuneita kieliopillisia morfeemeja on vain vähän. Esimerkkejä tällaisistakin kyllä löytyy: romania on lisännyt maskuliinin vokatiivin eli puhuttelusijan (joka esiintyi jo latinassa) rinnalle feminiinin vokatiivin, jonka pääte -*o* (esim. *soro* 'sisko!') on laina slaavilaisista kielistä (Feuillet 2012, 182).

Kreikan Traakiassa puhuttava romanikieli on lainannut turkista paitsi epäsuoran evidentiaalin (tiedon lähdettä ilmaisevan kieliopillisen muodon) myös sen merkitsimen -*miş* allomorfisessa asussa *muš*; romanikielessä se ei kuitenkaan ole sidonnainen vaan vapaa morfeemi ('kuulemma', 'muka', Adamou 2012, 24–25), joten sekään ei ole täydellinen esimerkki taivutusmorfeemin lainautumisesta. (Vrt. myös itäslaavista suomeen lainattuun johtomorfeemiin *-nik* > *-niekka*, joka suomessa on pikemminkin yhdyssadan jälkiosa kuin johdin, sillä se ei mukaudu vokaalisointuun, esim. *taitoniekka*, *kynäniekka*.)

Silloin kun kieliopillisen kategorian ilmaisu on analyyttinen eli perifrastinen (esimerkiksi syntaktisten suhteiden ilmaiseminen prepositioilla tai aikamuotojen ilmaiseminen apuverbillä), sen lainautuminen muistuttaa syntaktista lainautumista. Euroopan kielissä, varsinkin pitkin maanosan rannikkoalueita, esiintyy perifrastisia perfektitempuksia, joiden merkitys on varsin samankaltainen ja jotka muodostetaan partisiipista, johon liittyy apuverbi 'olla' (kuten suomessa) tai 'omistaa' (kuten ruotsissa; ks. Lindstedt 2000). Itse apuverbin äänneasu ei ole kielestä toiseen sama mutta perfektin muodostustapa on.

Tutkittaessa kielikontaktin vaikutuksia historiallisessa kielitieteessä keskeinen metodologinen kysymys on todistustaakka: vaativatko kontaktilähtöiset muutokset vai kielensisäiset muutokset erityistä todistelua, jos kumpikin selitys olisi periaatteessa mahdollinen. Perinteisesti kielentutkimuksessa on suosittu kielensisäisiä muutoksia nollahypoteesina, jota ei tarvitse erikseen todistaa, mutta sanojen lainautumista on sentään pidetty normaalina. Sellaisten sanaparien kuin ruotsin *gata* ja suomen *katu*, tai ruotsin *pojke* ja suomen *poika*, kohdalla lainautuminen on itsestään selvää, varsinkin kun ruotsin ja suomen välillä on paljon tällaisia sanapareja. Silloinkaan lähdekieli ja lainautumisen suunta ei ole aina itsestään selvä: suomen *katu* on lainaa ruotsista (mikä onkin ollut vallitseva lainaussuunta), mutta ruotsin *pojke* puolestaan on lainaa suomesta (ja sen samankaltaisuus englannin *boy*-sanan kanssa on siis pelkkää sattumaa).

Nykyään kieliopillisten piirteitten lainautumiseen sinänsä ei enää suhtauduta epäilevästi, mutta yleisesti voi sanoa, että mitä vaikeammin lainautuvista piirteistä on kysymys, sitä paremmin pitää yrittää kuvata

sitä sosiaalista tilannetta, jossa lainautuminen tai muu kontaktivaikutus on voinut tapahtua. Lainasanojen osalta riittää, että kielten kontakti on ylipäätään ollut mahdollinen – lainasanojen laatu voi sitten itsestään kertoa meille enemmän kontaktin luonteesta (ks. luku 3.3.). Sen sijaan esimerkiksi perfektitempuksen lainautumisen olettaminen edellyttää enemmän tietoa kontaktitilanteen laajuudesta ja pituudesta, kaksikielisten puhujien olemassaolosta ja kirjallisten traditioiden roolista yhteiskunnassa, varsinkin kun perifrastinen perfekti voi toki aina syntyä myös kielen sisäisten muutosten seurauksena.

Jos sama piirre esiintyy useammassa kuin kahdessa naapurikielessä, kontaktilähtöisyyden todennäköisyys kasvaa voimakkaasti, vaikkei piirteen lähdekieltä aina voisi osoittaa (vrt. luku 3.4. kieliliitoista). Euraasiassa on Balkanilta pitkälle Aasiaan ulottuva alue, jonka kielissä esiintyy tiedon lähdettä ilmaiseva evidentiaalisuuskategoria (Haarmann 1970; de Haan 2013). Alueen länsipään kieliin bulgariaan, makedoniaan ja albaniaan kategoria on selvästikin tullut turkista. Jonkinlaisista kieli-isänmaallisista syistä esimerkiksi osa bulgarialaisista lingvisteistä on halunnut pitää kielensä evidentiaalisuutta sisäsyntyisenä viitaten siihen, että evidentiaalisuus on kehittynyt myös moneen sellaiseen maailman kieleen, joka ei ole ollut missään yhteydessä turkkiin. Yksittäisen kielen kohdalla onkin aina mahdollista, että se on kehittänyt evidentiaalisuuden sisäisistä syistä, mutta on hyvin epätodennäköistä, että tällaiset kielet täysin sattumalta olisivat naapurikieliä, jotka muodostavat yhtenäisen alueen. Siksi kontaktiselitys on tässä tapauksessa uskottavampi.

## 3.2. Kontaktitilanteiden synkroninen tutkimus

Joskus kielikontakti-ilmiön asetelma tunnetaan hyvin ja siihen vaikuttavia tekijöitä voidaan havainnoida nykyhetkessä. Tällainen synkroninen kielikontaktien – tai usein pikemmin monikielisten käytänteiden – tutkimus on tyypillisesti kiinnostunut kontakti-ilmiöiden sekä erilaisten monikielisten tilanteiden analyysista sekä näistä tilanteista sosiaalisina ja yhteiskunnallisina ilmiöinä. Tällaisia tutkimusaloja ovat muun muassa akkommodaation ja koodinvaihdon (engl. *code switching*) tutkimus,

monikielisyystutkimus sekä kielenvaihdon (sa. *Sprachwechsel*, engl. *language shift*) tutkimus. Huolimatta eriytyneestä ja usein soveltavammasta tutkimusperinteestään toisen kielen oppimisen tutkimus tarkastelee niin ikään monen muun kielikontakti-ilmiön kannalta keskeisiä prosesseja.

Kaikkien edellä mainittujen tutkimusalojen piirissä on mahdollista tarkastella kohteena olevia ilmiöitä myös menneisyydessä, esimerkiksi toisen kielen oppimisen tutkimuksessa voidaan käyttää pitkittäisaineistoja kielitaidon karttumista mitattaessa. Diakronisesta kontaktitutkimuksesta nämä lähestymistavat kuitenkin tyypillisesti eroavat selittävien ja selitettävien tekijöiden osalta. Siinä missä esimerkiksi lainasanan eri murremuodot ja levikki tai toisesta kielestä lainatun rakenteen kieliopillinen rooli tunnetaan hyvin, lainautumiseen johtanut sosiolingvistinen asetelma ja muut tekijät ovat usein hämärän peitossa. Synkronisessa tutkimuksessa tarkastelun keskiössä ovat monikieliset käytänteet, ja tutkimusaineistoa on mahdollista kartuttaa haastatteluin ja jopa kokeellisesti, jolloin itse kontakti-ilmiöön vaikuttavia taustatekijöitä voidaan suoraan säädellä. Usein on kuitenkin kiistanalaista tai epäselvää, missä määrin havaitut synkroniset ilmiöt vaikuttavat kielenmuutokseen pidemmällä tähtäimellä.

Akkommodaatio tarkoittaa sitä, että puhujat modifioivat ilmaisuaan keskustelukumppanista riippuen. Koska akkommodaatiotutkimus tyypillisesti tarkastelee kahden puhujan vuorovaikutusta, on sen tutkimuksessa mahdollista soveltaa muun muassa kokeellisia menetelmiä (ks. Kokeellinen fonetiikka tk.). Tarkasteltavien selittävien tekijöiden osalta tutkimusasetelmat ovat usein samanlaisia kuin sosiolingvistiikassa (ks. Sosiolingvistiikka tk.). Perinteisesti akkommodaation tutkimus on keskittynyt äänteelliseen modifiointiin vuorovaikutustilanteessa samaa kieltä puhuvien kesken, mutta modifioinnin on todettu koskevan myös muita kielen osa-alueita ja olevan keskeistä myös ei-äidinkielisten ja äidinkielisten sekä kielen eri varieteettien puhujien välisessä kielellisessä vuorovaikutuksessa. Eräs esimerkki akkommodaatiosta on aikuisille kielenoppijoille suunnattu puhe (engl. *foreigner talk*), jonka roolia on pohdittu kieliliittotyyppisten ilmiöiden taustalla (ks. esim. Joseph 2010, 624–628). Peter Trudgill (1986) on puolestaan tarkastellut

akkommodaation merkitystä kielenmuutoksessa murrekontaktien osalta.

Koodinvaihto on monikielisille yhteisöille tyypillinen ilmiö, jossa puhuja käyttää elementtejä useammasta kielestä tai kielimuodosta saman puhetilanteen aikana. Koodinvaihdon tutkimus tarkastelee usein samantyyppisiä sosiolingvistisiä tekijöitä kuin akkommodaatiokin. Vaikka koodinvaihtoa on mahdollista tutkia myös vaikkapa kaunokirjallisessa korpuksessa, tyypillisesti tutkimuksessa käytetään kenttätyössä hankittua keskusteluaineistoa (ks. Kielten dokumentointi ja kieliopin kuvaus; Antropologinen lingvistiikka; Haastattelu aineistonkeruun metodina sosiolingvistiikassa tk.). Vaikka alalla ei vallitse yhtenäistä näkemystä koodinvaihdon ja sanaston lainautumisen suhteesta, on selvää, että molemmissa on kyse samankaltaisista prosesseista (ks. esim. Muysken 2000, 75–78).

Kielenvaihto eli tilanne, jossa puhujayhteisö siirtyy käyttämään toista kieltä kuin aiemmin, on jo varhain tunnettu kielisosiologinen asetelma – pohjautuvathan erilaiset substraattiselitykset juuri siihen. Käynnissä olevaa kielenvaihtoa alettiin kuitenkin tutkia puhujayhteisön ja yksilön kannalta vasta 1900-luvun loppupuoliskolla (ks. esim. tutkimuksen klassikko, Fishman 1964). Kielenvaihtoon liittyvä keskeinen käsite on *kielikuolema* (engl. *language death*, *language loss*; prosessiin viittaavana myös *language obsolescence*). Kielikuolema tarkoittaa tilannetta, jossa kieli ei enää välity sukupolvelta toiselle ja viimeisten puhujien kuollessa myös kieli katoaa. Kielikuoleman yhteydessä havainnoidaan usein käyttöalan kapenemisesta ja puhujayhteisön harvenemisesta johtuvia muutoksia kielessä (ks. Palosaari & Campbell 2011, 110–116). Samankaltaisia muutoksia tarkastellaan myös ensikielen attrition tutkimuksessa, jossa huomio on tyypillisesti yksilön ensikielen muutoksissa esimerkiksi muutettaessa alueelle, jossa valtakieli on jokin muu (ks. esim. Schmid 2011).

Kaksi- ja monikielisyyttä tutkitaan lukuisista eri näkökulmista (ks. esim. Edwards 1994). Varhaislapsuuden monikielisyyden tutkimuksen kysymyksenasettelut ja metodiikka ovat usein samoja kuin kielen omaksumisen (engl. *language acquisition*) tutkimuksessa. Sosiolingvistisessä monikielisyyden tutkimuksessa tarkastelun kohteena ovat monikieliset

yhteisöt ja yksilöt. Näiden lähestymistapojen piirissä tutkitaan usein kieli-identiteettiä ja monikielisyyttä resurssina sosiaalisessa vuorovaikutuksessa. Tällaisessa tarkastelussa kontaktilähtöiseen kielenmuutokseen ja esimerkiksi koodinvaihtoon liittyvä terminologia ja oletus diskreeteistä, selkeärajaisista kielimuodoista ovat saaneet osakseen kritiikkiä. Eritoten ilmiöiden diakroniaa tutkittaessa on aina tehtävä yleistyksiä kielistä, puhujayhteisöistä ja puhujista, mikä ei aina tee oikeutta esimerkiksi niille monille tekijöille, jotka vaikuttavat yksilön kielenkäyttöön. Tämä kritiikki on synnyttänyt esimerkiksi sellaisen käsitteen kuin *kieleily* (engl. *languaging*) kuvaamaan monikielisen yksilön toimintaa (kieleilyn ja kielikontaktin suhteesta ks. Lehtonen 2015, 304–307; ks. myös Lingvistinen etnografia toimintayhteisöissä tk.).

Diakroninen ja synkroninen kielikontaktien tutkimus täydentävät toisiaan: on oletettava, että kaikkien historiallisten kontaktimuutosten taustalla ovat perusluonteeltaan samankaltaiset monikieliset käytänteet, joita on mahdollista havaita nykyhetkessä. Tästä huolimatta diakronisen ja synkronisen tason yhdistäviä, uskottavia kontaktilähtöisen kielenmuutoksen malleja on olemassa varsin vähän. Mainita voi kuitenkin kaksi eritoten koodinvaihdon tutkimuksesta ammentavaa kontaktilähtöisen kielenmuutoksen mallia, Carol Myers-Scottonin (2002) ja Pieter Muyskenin (2013) mallit, joiden molempien keskiössä on kaksikielisen yksilön kielenkäyttö.

## 3.3. Kontaktien historian tarkastelu niiden tulosten valossa

Kielikontakteja voidaan joskus tutkia myös silloin, kun meillä ei ole juuri mitään historiallista tietoa kontaktitilanteesta. Paras esimerkki tästä on taas lainasanatutkimus: jos tietyt sanat pystytään äännehistoriallisin ja muin kielihistoriallisin argumentein osoittamaan lainasanoiksi, ne kertovat itsessään kontaktin tapahtuneen, ja sanojen merkitykset voivat tällöin kertoa paljon muinaisen kontaktin luonteesta. Ulkoisia ehtoja ei oikeastaan ole muita kuin se, että kyseisten kielten varhaisemmat muodot ovat voineet olla kosketuksissa keskenään eikä niitä ole esihistoriallisena aikana puhuttu esimerkiksi eri maanosissa. Kielten

puhuma-alueitten ei enää nykyään tarvitse olla kosketuksissa keskenään; esimerkiksi saamelaiskielissä on indoiraanisia lainasanoja.

Lainasanatutkimuksen ansiosta tiedämme suomalaisten kielellisten esi-isien muinaisista kontakteista ensin arjalaisten eli indoiraanisten kielten puhujien, sitten balttilaisten kielten puhujien ja vielä germaanisten kielten (ei vielä ruotsin) puhujien kanssa – näiltä kaikilta olemme saaneet paljon keskeistä kulttuurisanastoa, mutta koska nämä kontaktit ovat tapahtuneet tuhansien vuosien aikana ennen ajanlaskumme alkua, meillä ei olisi niistä juurikaan tietoa ilman näitä lainasanoja (ajoituksesta ks. esim. Kallio 2006b). Myöhemmistä suomalaisten kontakteista itäslaavien kanssa on saatu jo ainakin jotain historiallista tietoa, mutta tässäkin tapauksessa lainasanatutkimus voi täydentää historian antamaa kuvaa. Sellaiset slaavilaiset lainat kuin *akkuna* (> *ikkuna*), *lusikka, palttina, värttinä, turku* ja *suntio* vastaavat äänneasultaan vanhimmissa slaavilaisissa teksteissä esiintyviä sanoja ja ovat osittain jopa niitä arkaaisempia. Ne kertovat itämerensuomalais-slaavilaisten kontaktien alkaneen verrattain varhain, viimeistään ajanlaskumme ensimmäisen vuosituhannen puolessavälissä (ks. esim. Kallio 2006a), ja samalla ne antavat vahvistuksen monelle sellaiselle kantaslaavin äännepiirteelle, joka muuten jäisi kokonaan vertailevan rekonstruktion varaan (ks. Historiallinen ja vertaileva kielentutkimus tk.). Kaikki lainasanat kuvastavat myös aikansa materiaalista ja henkistä kulttuuria (ks. yleisesityksiä lainautumisesta: Häkkinen 1990; Haspelmath & Tadmor 2009).

Kieliopin muutos on huomattavasti vaikeampi osoittaa kontaktilähtöiseksi, jos meillä ei ole historiallista tietoa kontaktitilanteesta. Mikä tahansa kieliopin piirre on voinut syntyä kielen sisäisen kehityksen tuloksena, ja kieliopin piirteet eivät koskaan ole niin ainutkertaisia kuin sanojen äänneasu. (Useamman kuin kahden vierekkäisen kielen samankaltainen kielioppipiirre on jo todennäköisemmin kontaktilähtöinen, ja silloin puhumme kieliliitoista – ks. luku 3.4.) Toisin kuin lainasanat kielioppipiirre, joka on mahdollisesti lainautunut, ei myöskään kerro juuri mitään lainautumisajan kulttuurisen kontaktin laadusta, koska kieliopin rakenteella ja kulttuurilla ei ole suoraa suhdetta (vrt. kuitenkin luvut 3.5. ja 3.6. kontaktitilanteen ja kieliopillisen typologian suhteesta).

Jos halutaan antaa varoittava esimerkki siitä, mitä seuraa metodologisten vaatimusten höllentämisestä esihistoriallisten kielikontaktien tutkimuksessa, äärimmäinen tapaus on aikoinaan paljon julkisuutta saanut foneetikko Kalevi Wiikin (2002) teoria, jossa suuri osa eurooppalaisten kielten varhaisempaa historiaa selitettiin suomalais-ugrilaisen substraatin (ks. luku 2) avulla. Jos kaikki muutokset, jotka päällisin puolin tarkasteltuna voisivat olla kontaktilähtöisiä, myös oletetaan kontaktilähtöisiksi ilman mitään riippumatonta tukea, lopputuloksena on mielikuvituksellinen rakennelma, jolla on enää vähän yhtymäkohtia tunnettuihin faktoihin. Wiikin metodologisiin virheisiin kuului myös se, että hän surutta hylkäsi uniformitarianismin eli sen lähtökohdan, että menneisyydessä on oletettava vallinneen samojen säännönmukaisuuksien kuin nykyisyydessä, ellei meillä ole painavaa syytä olettaa muuta. Hänen mallissaan substraattivaikutus, lingua francan käyttö tai metsästäjä-keräilijöiden kielialueiden koko noudattivat menneisyydessä aivan eri lainalaisuuksia kuin nykyisin. (Wiikin teoksen kriittisiä arvioita on julkaistu useita, joista yksi on Lindstedt [2004].)

On kuitenkin olemassa lähestymistapa, jonka avulla naapurikielten kieliopillisia samankaltaisuuksia voidaan tarkastella jopa kaukaisen esihistorian näkökulmasta ilman, että sorrutaan perusteettomaan spekulointiin. Käänteentekevässä monografiassaan *Linguistic Diversity in Space and Time* Johanna Nichols (1992) osoitti, että monen kieliopillisen piirteen levinneisyys maailman kielissä korreloi kielten sijainnin kanssa koko maailman mittakaavassa. Monet piirteet muodostavat länsi-itäsuuntaisia kliinejä, toisin sanoen piirrettä kuvaava indeksi muuttuu paikan monotonisena funktiona. ”Länsi” ja ”itä” määräytyvät tässä maapallon asutushistorian mukaan: länttä ovat Afrikka, josta ihmislajin ekspansio alkoi, Lähi-itä ja Eurooppa, itää ovat puolestaan Australia ja Etelä-Amerikka, jotka olivat ekspansion pääteasemia, ja niiden välissä ovat Tyynenmeren rannikko ja Pohjois-Amerikka. Nicholsin selitys tälle on se, että itäiset piirteet olivat itäisempien kielten piirteitä jo kymmeniä tuhansia vuosia sitten, kun ekspansio alkoi, eli ne edustavat laajenevan ihmiskunnan eturintaman piirteitä, kun taas läntisemmät piirteet olivat leimallisia myöhemmin liikkeelle lähteneille ihmisryhmille (Nichols 1992, 254–259). Toki kielet ovat omaksuneet uusia alueittain

samankaltaisia piirteitä ihmiskunnan levittäytymisen jälkeenkin, mutta tällainen alueellinen konvergenssi ei olisi voinut tuottaa koko maailman mittaisia kliinejä.

Nichols (1992) liikkuu sellaisessa aikasyvyydessä, johon historiallis-vertaileva menetelmä ei ulotu (ks. Historiallinen ja vertaileva kielentutkimus tk.). Niinpä meillä ei ole mitään keinoa tietää, ovatko hänen löytämänsä vanhimmat alueelliset samankaltaisuudet kielisukulaisuuden vai kielikontaktin aiheuttamia – todennäköisesti kyse on kummastakin, mutta emme voi mitenkään tietää, missä suhteessa. Siitä huolimatta hän on saanut varsin vakuuttavia tuloksia. Vaikka kielikontaktien tutkimus siis edellyttää areaalisuuden eli maantieteen huomioon ottamista, areaalisuutta voidaan tutkia olettamatta mitään kielikontakteista. Seuraavassa kappaleessa palaamme kuitenkin sen verran pieniin aikasyvyyksiin, että areaalisuus ja kontaktit ovat liitettävissä toisiinsa.

## 3.4. Areaalisuus: kieliliitot ja kontaktialueet

Useimmiten kielikontaktien tutkimus selvittelee sitä, miten jokin piirre on lainautunut yhdestä kielestä toiseen, tai synkronisesti, miten jonkin kielen elementit käyttäytyvät ja mitä niillä tehdään puhuttaessa toista kieltä. Mikäli lainanantajakieli on luonteeltaan superstraatti eli prestiisiltään korkeampi varieteetti, lainautuminen on todennäköisesti yksisuuntaista eli toinen kieli on lainojen lähde ja toinen vastaanottaja (ks. kuitenkin luku 3.6. L2-puhujien vaikutuksesta). Jos kielten välillä vallitsee kuitenkin adstraattisuhde ja kumpikaan kieli ei merkittävästi dominoi, on varsin luonnollinen kysymys, mitä tapahtuu lainanantajakielelle itselleen. Mikäli lainautumisen taustalla on kahden erikielisen yhteisön vuorovaikutus, jossa osa molempien kielten puhujista osaa ainakin jossain määrin myös toista kieltä, voi lainautumista tapahtua molempiin suuntiin.

Kuten luvussa 2 todettiin, kielikontaktien tutkimuksen syntyyn liittyvät keskeisesti havainnot tietyillä alueilla puhutuista kielistä, joilla on paljon yhteisiä piirteitä, jotka eivät selity kielten sukulaisuudella. Tällaisten alueiden nimitykseksi vakiintui 1900-luvun puolivälissä *kieliliitto*, ja

havaintoja tämän tyyppisistä kielellisistä alueista on tuon jälkeen tehty lukuisia. Tyypillistä näille havainnoille on se, että yhteisten piirteiden alkuperää ei tunneta, eli jaetut piirteet eivät selity esimerkiksi yhteisellä superstraatilla, mikä sekin voi kyllä johtaa samankaltaiseen lopputulokseen. Ajatellaan esimerkiksi Venäjän federaation vähemmistökieliä: lähes kaikki niiden puhujat osaavat myös venäjää, ja venäjästä voi näin ollen lainautua sama rakenne useampaan kieleen niin, että lainattujen piirteiden osalta vähemmistökielet saattavat alkaa muistuttaa toisiaan. Tällaista superstraattilähtöistä alueellista samankaltaisuutta ei kuitenkaan yleensä kutsuta kieliliitoksi.

Eräs ainakin teoreettinen mahdollisuus, joka voisi johtaa niin ikään kielten samankaltaistumiseen tietyllä alueella, on yhteinen substraatti. Mikäli useamman kielen historiassa on tapahtunut kielenvaihtoa ja taustalla on kaikissa sama substraattikieli, voivat muutokset olla samankaltaisia. Tällaisten selitysten ongelma on usein se, että niitä on melkein mahdotonta osoittaa oikeiksi kiistattomasti. Substraattiselitysten aiemman suosion taustalla lienee se, että substraatti oli varhaisimpia tunnettuja kielikontaktin tyyppejä ja muita mekanismeja tunnettiin huonosti. Ei ole kuitenkaan poissuljettua, etteikö jonkin tunnetun kieliliittoilmiön taustalla olisi ainakin osittain yhteinen substraatti. Esimerkiksi manner-Kaakkois-Aasian kieliliiton äänteelliset piirteet voivat hyvin selittyä mon-khmer-kielten (esim. vietnam) substraatilla (ko. kieliliitosta ks. Enfield 2005). Samaten Intian kieliliitolle luonteenomaiset retrofleksikonsonantit voivat olla perua dravidakielten substraattivaikutuksesta indoarjalaisissa kielissä (Intian kieliliitosta ks. Emeneau 1956).

Jos siis kieliliiton taustalla on yleensä ristiinlainautuminen prestiisiltään suurin piirtein samanlaisten kielten välillä, onko kyseessä lopulta vain alue, jonka kielissä sattuu olemaan lainoja useammasta kielestä? Tällaista näkemystä edustaa esimerkiksi Lyle Campbell (2006), joka olisi valmis hylkäämään *kieliliiton* käsitteen ja puhumaan ainoastaan yksittäisten paikallisten lainojen kumuloitumisesta tietyllä alueella. Hieman samassa hengessä Bernd Heine ja Tania Kuteva (2006, 286–287) asettuvat vastustamaan kieliliittojen osalta usein käytettyä *konvergenssin* käsitettä, jolla kuvataan tilannetta, jossa kieliliiton kielten rakenteet lähentyvät toisiaan. Myös Heine ja Kuteva käsittelevät kieliopillisia

kontakti-ilmiöitä yksittäisten lainautumisprosessien kautta, ja heidän luomassaan mallissa on aina selkeästi tunnistettava lainan antaja ja vastaanottaja.

Vaikka edellä mainittu kritiikki hyväksyttäisiin, tulee kuitenkin selittää, miksi joillain alueilla tällaista monensuuntaista lainautumista tapahtuu ja miksi toisilla taas ei. Lisäksi herää kysymys siitä, miksi esimerkiksi Balkanilla, jossa useamman alueen kielen historia tunnetaan melko hyvin, on kuitenkin usein mahdotonta määritellä jaetun piirteen alkuperää. Eräs tarjottu selitys on, että kieliliitto olisi sittenkin enemmän kuin yksittäisten lainautumisten summa: Jaetuilla piirteillä ei välttämättä ole Campbellin (2006) sekä Heinen ja Kutevan (2006) olettamaa selkeää lähdettä, vaan kieliliiton taustalla oleva sosiolingvistinen asetelma suosii ja levittää tietyntyyppisiä rakenteita. Balkanin kieliliiton osalta monet jaetut piirteet ovat analyyttisiä rakenteita, jotka ovat korvanneet kaikissa kielissä aiemmat synteettiset rakenteet (Lindstedt 2019; Balkanin kieliliitosta suomeksi ks. Lindstedt 2009a, jossa kartan selitykset ovat kuitenkin vaihtuneet keskenään; ks. myös luku 3.6.).

Kieliliittojen tutkimus on osa niin kutsuttua areaalilingvistiikkaa, joka tarkastelee kielellisten piirteiden alueellista levikkiä. Useimmiten kieliliittojen tutkimuksessa on taustalla jonkinlainen historiallinen hypoteesi niiden syntyyn vaikuttaneista tekijöistä. Usein kuitenkin havaitaan, että jollakin tietyllä laajallakin, esimerkiksi mantereen kokoisella maantieteellisellä alueella kielet suosivat tietyntyyppisiä rakenteita. Tästä seuraa, että kielen puhuja-alueen maantieteellinen sijainti voi olla yhtä hyvä kielen rakenteen selittäjä kuin kielen kuuluminen johonkin kielikuntaan. Lauri Hakulinen (1968, 388–392) kiinnittää huomion siihen, että suomen kieli muistuttaa paljon muita eurooppalaisia kieliä sanastonsa rakenteeltaan johtuen lukuisista käännöslainoista. Östen Dahl (2008) laajentaa tarkastelun myös kielioppiin ja toteaa, että suomi on rakennepiirteiltään varsin eurooppalainen kieli verrattuna esimerkiksi idempänä puhuttuihin sukukieliinsä.

Suomen tapauksessa on selvää, että kielen rakenteiden osalta samankaltaisuuden täytyy olla seurausta kielikontakteista. Tämän tyyppisessä areaalilingvistisessä tarkastelussa kielikontaktit ovat kuitenkin vain yksi – joskin varteenotettava – kielipiirteiden alueellisen jakauman selittäjä.

Kun tarkastellaan jonkin piirteen levikkiä ja aineistona on edustava otos maailman kielistä, ei voida poissulkea sitä, etteikö piirteen edustumien taustalla olisi myös muinainen kielisukulaisuus. Tällaiset typologiset piirteet eivät kuitenkaan itsessään kelpaa todistusaineistoksi kielisukulaisuudesta juuri sen vuoksi, että ne voivat levitä myös kontaktissa, ja kielisukulaisuuden todistamiseen tarvitaan aina historiallis-vertailevan kielitieteen menetelmiä säännöllisten äännevastaavuuksien löytämiseksi (ks. myös luku 3.3.; Historiallinen ja vertaileva kielentutkimus tk.).

Havainnot kielipiirteiden alueellisesti merkitsevästä levikistä ovat johtaneet siihen, että määrällinen kielitypologinen tutkimus pyrkii aina huomiomaan areaalisia tekijöitä tarkastellessaan kielen universaaleja (ks. Kielitypologian menetelmät tk.). Tähän tarkoitukseen käytetään erilaisia otoksia, joissa maapallo on jaettu eri kokoisiin makroalueisiin (menetelmästä ks. Dryer 1989). Balkanin kaltaisten kieliliittojen tutkimuksen rinnalla mutta osittain myös siitä riippumatta on kehittynyt kooltaan huomattavasti suurempien makroalueiden tutkimus. Toisin kuin kieliliittojen makroalueiden synty on useimmiten sellaisessa aikasyvyydessä, että ei voida saavuttaa varmuutta siitä, juontuvatko jaetut piirteet kielikontakteista vai yhteisestä muinaisesta alkuperästä (ks. luku 3.3.). Pääasiassa kontaktioletuksiin perustuvat makrotason mallit kilpailevatkin usein selityksinä omnikomparativististen etäkielikuntien kanssa. Näillä tarkoitetaan kielikuntahypoteeseja, jotka perustuvat kielten pinnallisiin äänteellisiin samankaltaisuuksiin (ks. myös Historiallinen ja vertaileva kielentutkimus tk.).

## 3.5. Kontaktikielet

Kielten välinen kontakti voi johtaa siihen, että syntyy uusia kieliä, jotka eivät ole kontaktitilanteeseen osallistuneiden kielten tytärkieliä käsitteen perinteisessä merkityksessä (ks. Historiallinen ja vertaileva kielentutkimus tk.). Tällaisia uusia kieliä kutsutaan *kontaktikieliksi* – tosin termillä tarkoitetaan joskus mitä tahansa erikielisten välisessä kontaktissa käytettävää kieltä, mutta suppeammassa merkityksessä kontaktikielet eivät ole pelkästään käytössä kontaktitilanteessa, vaan ne ovat myös

syntyneet siinä. Tällaisten kontaktikielten kolme päätyyppiä ovat *pidginit, kreolit* (kreolikielet) ja *sekakielet*. Niiden tutkimus on itsenäinen lingvistiikan ala, jolla on omat vakiintuneet julkaisukanavansa (esimerkiksi *Journal of Pidgin and Creole Languages* ja kirjasarja *Creole Language Library*, uudemmalta nimeltään *Contact Language Library*). Johdantotekstejä kontaktikielten tutkimukseen ovat esimerkiksi Bartens (2009), Thomason (2001, 157–221) ja Arends, Muysken ja Smith (1995). Laajempi kokoomateos on Kouwenberg ja Singler (2008).

Pidginit eivät ole kenenkään äidinkieliä. Ne ovat syntyneet kahden tai useamman kielen puhujien kontaktitilanteessa, jossa eri kielten puhujat eivät ole osanneet toistensa kieltä mutta ovat joutuneet kommunikoimaan keskenään esimerkiksi kaupankäynnin vuoksi. Vastoin yleistä käsitystä pidginit eivät ole ”sekakieliä”, jotka olisivat syntyneet kahden tai useamman kielen yhteensulautumisesta. Niiden sanasto on useimmiten pääosin peräisin yhdestä, prestiisiltään korkeammasta kielestä, jota kutsutaan *leksifioijaksi*. Pidginien kielioppi puolestaan ei pohjaudu suoraan mihinkään muuhun kieleen vaan pikemminkin kuvastaa kommunikaation yksinkertaisia perustarpeita. Konkreettisessa kommunikaatiotilanteessa puhujien omat äidinkielet voivat toki vaikuttaa puhunnosten muotoon, esimerkiksi sanajärjestykseen. Morfologiaa pidgineissä on hyvin vähän tai ei ollenkaan.

Yksi tunnetuimmista pidgineistä oli lingua franca -niminen kieli, jota käytettiin Välimeren alueella ristiretkien ajalta 1800-luvulle asti (tätä ei pidä sekoittaa *lingua franca* -termin uudempaan käyttöön merkityksessä ’yhteiskieli’, esimerkiksi ”Englanti lingua francana”, ks. luku 2.). Sen leksifioijana olivat romaaniset kielet, erityisesti italia. Suomea lähin vakiintunut pidginkieli oli russenorsk, jota norjalaiset ja venäläiset kalastajat ja kauppiaat käyttivät kaupankäynnissään Jäämeren alueella 1700-luvun lopulta 1900-luvun alkuun. Russenorsk oli siinä mielessä poikkeuksellinen pidgin, että sillä oli kaksi pääleksifioijaa, venäjä ja norja; sen lisäksi siinä oli esimerkiksi englantilaisia ja hollantilaisia sanoja, jotka merenkulkijoitten mukana ovat levinneet moneen muuhunkin maailman pidginiin. Russenorskiksi käyty keskustelu aloitettiin usein lauseella *moja po tvoja* ’minä puhun sinun kieltäsi’, mikä viittaisi siihen, että norjalaisten mielestä russernorsk oli (eräänlaista) venäjää ja venäläisten

mielestä (eräänlaista) norjaa. Täydellisin lähde russenorskista on Broch ja Jahr (1984).

Jotkin pidginit ovat saaneet äidinkielisiä puhujia; näin kävi erityisesti tietyissä siirtomaaolosuhteissa, esimerkiksi viljelyksillä, joilla työskenteli eri seuduilta kotoisin olevia orjia. Kun pidgin saa äidinkielisiä puhujia, sen käyttöala laajenee ja rakenne täydellistyy muitten äidinkielenä käytettävien ihmiskielten mukaiseksi, jolloin siitä tulee *kreoli*. Kreolikielellä on siis äidinkielisiä puhujia, mutta joskus aikaisemmin sillä ei ole ollut. Historiallisen kielentutkimuksen kannalta sanotaan, että sen historiassa on ollut sukupolvien välisen siirron katko (engl. *broken transmission*) eikä se asetu minkään kieliperheen sukupuuhun samassa mielessä kuin katkeamatta siirtyneet kielet. Historiallis-vertailevaa menetelmää (ks. Historiallinen ja vertaileva kielentutkimus tk.) ei siis voi suoraan soveltaa kreolien historian tutkimukseen.

Tunnettu esimerkki kreoleista on haitinkreoli, jonka leksifioija on ranska (ja jota ei siis pidä sekoittaa Haitilla puhuttavaan ranskan varieteettiin). Alankomaihin kuuluvilla Karibianmeren saarilla puolestaan puhutaan papiamentoa, jonka pääasiallinen leksifioija on portugali, mutta jossa on myös paljon espanja- ja hollantiperäistä sanastoa. Suomalaisten päiväkotien sormileikkilaulua ”Täällä liikkuu viisi sormea” (san. Soili Perkiö) lauletaan usein myös eksoottisella vieraalla kielellä (*sinku dede mi tin na mi man* ’viisi sormea minulla on kädessäni’), mutta juuri kukaan ei Suomessa tunnista kieltä papiamentoksi.

Englantilaispohjaisia eli englannin leksifioimia kreoleja ovat esimerkiksi tok pisin, joka on yksi Papua-Uusi-Guinean virallisista kielistä, ja Länsi-Afrikan pidginenglanti, jolle BBC vastikään avasi oman uutiskanavan[1]. Kumpikaan kieli ei siis nimestään huolimatta ole enää pidgin vaan kreoli.

Jotkut tutkijat ovat sitä mieltä, ettei kreolin synty aina edellytä edeltävää pidginiä. Thomason ja Kaufman (1988, 147–166) ovat esittäneet, että esimerkiksi erikielisten orjien keskuudessa saattoi tapahtua ”äkillinen kreoloituminen”: heti ensimmäisten orjille syntyneiden lasten (yhdeksi) äidinkieleksi tuli heidän vanhempiensa puutteellisesti oppima siirtomaaisäntien kieli, ja kreoli syntyi ilman edeltävää pidginvaihetta. Ainakin osittain tämä on myös määrittelykysymys, sillä irrallisista yrityksistä

kommunikoida huonosti osatulla kielellä on jatkumo rakenteeltaan jossakin määrin vakiintuneeseen pidginiin. Varmaankaan kreolin syntyminen ei ole aina edellyttänyt vakiintunutta pidginiä.

On myös mainittava se käsitys, jonka puolesta on argumentoinut esimerkiksi Salikoko Mufwene (2008, 115–201), että kreolit eivät muodosta mitään poikkeuksellista ryhmää ihmiskielten joukossa. Hänen mukaansa latinan muuntuminen ranskaksi kielenvaihdon kautta ja kielikontaktien vaikutuksen alaisena oli aivan samankaltainen prosessi kuin ranskan muuntuminen haitinkreoliksi, joka siis kuuluu romaanisten kielten sukupuuhun yhtenä normaalina haarana. Kreolikielet on eroteltu omaksi ryhmäkseen vain niiden puhujien alhaisen yhteiskunnallisen statuksen vuoksi. Mufwenen ajatukset ovat terveellinen muistutus siitä, että kontaktilingvistiikan ilmiöt eivät voi olla täysin erilaisia kreoleiden synnyssä kuin muitten ihmiskielten kohdalla. Toisaalta Mufwene ei kuitenkaan ota kantaa siihen periaatteelliseen metodologiseen eroon, ettei kreoleita voi käyttää historialliseen rekonstruktioon (esim. tok pisiniä kantagermaanin rekonstruktioon) samalla tavalla kuin muita kieliä (Thomason & Kaufman 1988, 200–201). Niiden historiassa on siis ollut poikkeuksellisen uudelleenjärjestäytymisen vaihe, vaikkeivät kaikki tutkijat halua kutsua sitä *katkenneeksi siirroksi*. McWhorter (2018) argumentoi vahvasti sen puolesta, että kreolit ovat kuin ovatkin alkuperältään ainutlaatuisia ja itse asiassa ainoita aidosti uusia ihmiskieliä.

Yleisen kontaktilingvistiikan kannalta on tässä yhteydessä hyvä varoittaa muutamasta termistä. *Kreoloituminen* tarkoittaa pidginin kohdalla sitä, että se nativisoituu eli saa äidinkielisiä puhujia ja alkaa heidän puheyhteisössään muuttua. Kreoloitumiseen liittyy siis rakenteen ja funktioiden kehittyminen ja monimutkaistuminen. Mutta jos asiaa tarkastellaan leksifioijakielen kannalta, ”kreoloituminen” saattaakin päin vastoin tarkoittaa rakenteen yksinkertaistumista, koska kreoleilla on yksinkertaisempi kielioppi kuin leksifioijalla (tosin tästäkin on eri käsityksiä, ks. luku 3.6.). Varsinkin Thomasonin ja Kaufmanin (1988) esittämä ”äkillinen kreoloituminen” olisi ennemmin yksinkertaistumista kuin monimutkaistumista. *Dekreoloituminen* taas tarkoittaa sitä, että kreoli alkaa rakenteeltaan lähestyä leksifioijaa ja lainaa siltä sanastoa ja kielioppia, koska sillä on suurempi prestiisi.

Sen sijaan on syytä välttää *kreoloituminen*-termin käyttöä kuvaamaan mitä tahansa kielen rakenteen radikaalia muutosta kontaktitilanteessa, jos tuloksena ei ole kreolikieli. Esimerkiksi afrikaansia ei pidä kutsua ”kreoloituneeksi hollanniksi”, koska kielen siirto sukupolvelta toiselle ei hollannin ja afrikaansin välissä ole missään vaiheessa katkennut. Jos halutaan korostaa ei-äidinkielisten puhujien (todennäköisesti vahvaa) roolia sen kehityksessä (Thomason & Kaufman 1988, 251–256), tarjolla on sellaisia termejä kuin *kreoloidi* ja *semikreoli*. Tämäkin liittyy seuraavassa luvussa 3.6. käsiteltävään kysymykseen ei-äidinkielisten puhujien merkityksestä kielen typologisen muutoksen kannalta.

Termit *superstraatti* ja *substraatti* ovat kreolistiikassa käytössä hieman eri merkityksessä kuin muussa kontaktilingvistiikassa. *Superstraatilla* tarkoitetaan useimmiten leksifioijaa tai muuta kreoliin vaikuttavaa, prestiisiltään korkeaa kieltä. *Substraatti* taas tarkoittaa niitä kieliä, jotka olivat kreolia edeltävän pidginin puhujien äidinkieliä. Yksi keskeinen kysymys kreolistiikassa on, missä määrin kreolien kielioppi on peräisin superstraatista, substraatista tai kielellisistä universaaleista.

Termiä *sekakieli* (engl. *mixed language*) käytetään kontaktilingvistiikassa paljon vähemmän kuin yleisessä kielenkäytössä. Se, että jokin kieli on lainannut paljon sanastoa toiselta, ei tee siitä sekakieltä: englanti on yhä germaaninen kieli, vaikka se on lainannut paljon ranskasta, jiddiš on myös germaaninen kieli slaavilaisista ja heprealaisista lainoistaan huolimatta, ja albania on yhä oma haaransa indoeurooppalaisessa kieliperheessä, vaikka sen suoraan kantakielestä peritty sanasto on supistunut pieneksi romaanisten ja slaavilaisten lainojen tieltä.

Kansanomaisissa esityksissä pidginit ja kreolit selitetään joskus kahden tai useamman kielen sekoituksiksi, mitä ne siis missään tapauksessa eivät ole. Maailmassa on kuitenkin muutamia todellisia sekakieliä, jotka muodostavat pidginien ja kreolien ohella kontaktikielten kolmannen ryhmän (Thomason 2001, 196–221). Ne eivät ole syntyneet erikielisten väestöjen kohtaamisessa vaan yhden väestöryhmän keskuudessa kahden kielen kohdatessa. Sen sijaan, että ryhmä hylkäisi alkuperäisen kielensä valtakielen levitessä heidän alueelleen, se luokin uuden kielen, jossa on elementtejä kummastakin.

Sekakielille on tyypillistä, että kielen pääkomponentit, esimerkiksi sanasto, substantiivien morfologia tai verbien morfologia, voivat olla selvästi eri lähteistä. Esimerkiksi Mednyisaaren aleutti, jota puhuttiin 1800-luvulla ja vielä 1900-luvullakin pienellä Tyynenmeren saarella Kamtšatkan itäpuolella, oli muuten sanastoltaan ja kieliopiltaan aleuttia, mutta aleutin monimutkainen finiittiverbien morfologia oli korvattu venäjän mukaisella verbitaivutuksella. Ecuadorissa puhuttavan media lengua -kielen kielioppi puolestaan pohjautuu ketšuaan mutta sanasto espanjaan; puhujat muodostavat oman ryhmänsä, jotka eivät identifioidu ketšuan tai espanjan puhujiksi. Myös nk. pararomanikieliä, joissa romanikielen sanasto liittyy paikallisen kielen (esimerkiksi englannin tai suomen) kielioppiin (Matras 2002, 242–248), voi pitää sekakielinä. Tässäkin vain on ongelmana, että määrittelystä riippumatta termi *sekakieli* tuo puhujille itselleen ja laajemmalle yleisölle aina negatiivisia assosiaatioita.

Sekakielten tutkimuksen kannalta mielenkiintoinen kielimuoto on helsinkiläinen stadin slangi – nimenomaan sen alkuperäinen, 1800- ja 1900-lukujen vaihteessa puhuttu versio, jossa kielioppi oli suomen mukainen mutta sanasto pääosin ruotsista (Kallio 2007). Se ei ollut yllä kuvatun kaltainen sekakieli siinä mielessä, että sillä olisi ollut äidinkielisiä puhujia, mutta se oli kuitenkin tietyn ryhmän identiteetin merkitsin. Toisaalta se ei oikein muistuta pidginiäkään, koska sillä oli monimutkainen, yhdestä kielestä lainattu kielioppi. Joka tapauksessa sitä voidaan pitää suomen ja ruotsin kontaktista syntyneenä varieteettina – kummallakin kielellä oli tuohon aikaan suunnilleen yhtä paljon puhujia Helsingissä (Kallio 2007, 176) – mutta on määrittelykysymys, oliko se uusi kontaktikieli vaiko vain suomen erikoinen kontaktivarieteetti (eli varieteetti, jossa kontaktivaikutus on huomattavan voimakas saman kielen muihin varieteetteihin verrattuna). Koska Kallio pitää kieliopillisia elementtejä kielisukulaisuuden määräävänä tekijänä, hän pitää stadin slangia suomalais-ugrilaisten kielten sukupuuhun kuuluvana. Tällä perusteella media lenguakin olisi ehkä ketšuan sukukieli, ei sekakieli, mutta Mednyisaaren aleutin luokittelu aleutin tai venäjän sukukieleksi olisi jo mielivaltaista, koska jako kulkee keskeltä kielioppia.

Lindstedt (2009b) on esittänyt, että kontaktikielten tutkimuksen metodeja voidaan käyttää jopa esperanton alkuperän, muutoksen ja

äidinkielistymisen tutkimuksessa. Esperanton peruskielioppi ja -sanasto suunniteltiin tietoisesti, mutta esperantonkielisissä teksteissä ja sen puhujien kielenkäytössä on runsaasti ilmiöitä, jotka eivät kuvasta julkaistuja suppeita kielioppeja vaan ovat sen kieliympäristön vaikutusta, jossa kieltä on käytetty. Esperantoa voidaan tässä mielessä pitää kontaktikielenä, jossa on romaanis-germaaninen superstraatti ja slaavilainen substraatti (kreolistiikassa käytetyssä merkityksessä, ks. ylh.). Äidinkieliset esperanton puhujat ovat jossain määrin samassa tilanteessa kuin pidginiä kreoloivat varhaiset puhujat.

Kielikontaktien tutkimuksen kannalta tärkein kontaktikielten antama opetus on, ettei ihmiskielessä ole mitään osaa, johon tietynlainen kontaktitilanne ei voisi vaikuttaa radikaalistikin. Kreolistiikan piirissä syntyneillä tavoilla tarkastella kielten sosiohistoriallista taustaa ja kielen siirtoa sukupolvelta toiselle on paljon annettavaa myös sellaisten varieteettien tutkimukselle kuin afrikaans tai stadin slangi, jotka eivät ainakaan ilmeisen selvästi ole itse kontaktikieliä tämän osan alussa kuvatussa suppeammassa mielessä.

## 3.6. Kielikontaktien vaikutus kielen tyyppiin

Kontaktit eivät tuo kieleen pelkästään yksittäisiä toisista kielistä lainattuja piirteitä, vaan ne voivat vaikuttaa myös kielen koko rakenteeseen. Tämä liittää kontaktien tutkimuksen kielitypologiaan ja sen piirissä käytyyn keskusteluun kielten kompleksisuudesta eli monimutkaisuudesta ja sen vähenemisestä eli yksinkertaistumisesta.

Strukturalistisessa kielentutkimuksessa oli pitkään vallalla aksiooma, jonka mukaan kaikki kielet ovat yhtä monimutkaisia (ks. Kusters 2003, 1–5). Tämä oli aikanaan ymmärrettävä reaktio niihin varhaisempiin etnosentrisiin käsityksiin, joiden mukaan useimmat kielet ovat jotenkin primitiivisempiä kuin indoeurooppalaiset ja mahdollisesti seemiläiset kielet (vaikka jo Darwin tiesi tämän käsityksen vääräksi, vrt. ”[t]he perfectly regular and wonderfully complex construction of the languages of many barbarous nations”; Darwin 1874, luku III, jossa viitataan kielitypologian pioneeriin Friedrich Schlegeliin). Koska kielihistoriassa on

kuitenkin helppo löytää esimerkkejä kielen jonkin osan ilmeisestä yksinkertaistumisesta (esimerkiksi latinan monimutkaisen taivutusmorfologian yksinkertaistumisesta ja kadosta romaanisissa kielissä), yhtäläisen kompleksisuuden aksioomaa oli täydennettävä sillä oletuksella, että yksinkertaistavan muutoksen jotenkin kompensoi kielen jonkin toisen komponentin monimutkaistuminen. Tällaiselle oletukselle ei kuitenkaan ole löytynyt empiiristä tukea, eikä kukaan myöskään ole voinut esittää, millainen koko kielen kattava säätelymekanismi pitäisi huolen tällaisesta kompleksisuuden kokonaishallinnasta (Trudgill 2011, 15–16). Kokonaiskompleksisuuden mittaaminen olisi sitä paitsi vaikeaa, koska emme tiedä, mitkä kaikki piirteet pitäisi ottaa huomioon tai miten asettaa kielen eri tasoja kuvaavat mittarit samalle asteikolle (Miestamo 2017, 238).

Nykyään on voittamassa alaa se käsitys, että on mielekästä puhua eri kielten kompleksisuuden eroista, kunhan määritellään, millaista kompleksisuutta tarkoitetaan. Dahl (2004, 19–55), McWhorter (2011, 1–2) ja Trudgill (2011, 15–24) ovat analysoineet *kompleksisuus*-käsitteen eri puolia. Usein esitettyjä parametreja ovat 1) sääntöjen monimutkaisuus, eli kuinka laaja on kieliopin jonkin osan tyhjentävä kuvaus, 2) läpinäkyvyyden puute kieliopissa tai leksikossa, eli kuinka kaukana ollaan ”yksi merkitys – yksi muoto” -periaatteesta (vrt. Anttila 1972, 100–102, passim), ja 3) redundanssi, eli vaatiiko kielioppi merkitsemään seikkoja, jotka selviävät kontekstista (paradigmaattinen redundanssi) tai jotka on jo merkitty toisessa kohtaa lausetta (syntagmaattinen redundanssi). Redundanssi lisää ennen kaikkea kielen puhujan prosessointitaakkaa, mutta kuulijan kannalta se on viestin ymmärtämistä helpottava. Miestamo (2008; 2017) on esittänyt, että kompleksisuuden mittaaminen on redusoitavissa kahteen parametriin: kuinka monia semanttisia tai pragmaattisia erotteluja kielessä merkitään, ja kuinka kaukana ollaan ”yksi merkitys – yksi muoto” -periaatteesta.

Kompleksisuus korreloi varsin pitkälti sen kanssa, kuinka helppoa kieltä on oppia vieraana kielenä kriittisen kielenomaksumisiän jälkeen, mutta tämä *L2-vaikeus* on kuitenkin erillinen, empiirisesti tutkittavissa oleva suure, joka on ainakin osittain sidoksissa oppijan ensikielen rakenteeseen. Jos kielessä on foneemisia tooneja (kuten kiinassa) tai

tonaalisia aksentteja (kuten ruotsinruotsissa), ne eivät välttämättä tee kielen fonologian kuvauksesta kovin paljon monimutkaisempaa mutta saattavat silti olla huomattavan vaikeita L2-oppijan kannalta.

Edellisessä luvussa 3.5. käsiteltyjen pidginien ja kreolien tutkimuksessa yksi keskeinen kysymys on ollut, ovatko tällaiset kontaktikielet rakenteeltaan jossakin mielessä yksinkertaisempia kuin muut ihmiskielet. Pidginien osalta kysymys ei ole kovinkaan kiistanalainen, koska ne eivät ole kenenkään äidinkieliä ja niiden käyttöala on rajattu, joten niiden rakenteellinen yksinkertaisuus on odotuksenmukaista. Mutta kun McWhorter (2001) esitti, että kreolikielten kieliopit ovat yksinkertaisempia kuin muitten äidinkielinä puhuttujen kielten, monet kielentutkijat olivat jyrkästi eri mieltä (jo samaan *Linguistic Typology* -lehden numeroon oli kerätty monia kommenttipuheenvuoroja). Vastaväitteiden takana oli selvästikin pelko, että yhteiskunnalliselta arvostukseltaan alhaisten kreolikielten nimittäminen yksinkertaisiksi estäisi niiden emansipaatiota ja olisi jopa rasistista.

McWhorterin teos (2011) kielten yksinkertaisuudesta ja monimutkaisuudesta perustelee hänen argumenttejaan laajemmin. Hänen perusteesinsä, jonka mukaan kielten kompleksisuuserot riippuvat merkittävästi siitä, mikä rooli toisen kielen oppimisella ja siis L2-puhujilla on ollut niiden historiassa, on sovellettavissa muihinkin kieliin kuin kreoleihin, vaikka kreolit ovat tietysti tässä mielessä selvin tapaus. Kustersin (2003) uraauurtava väitöskirja osoitti, kuinka arabian, skandinaavisten kielten, ketšuan ja suahilin tiettyjen varieteettien verbintaivutus yksinkertaistui niiden levitessä nopeasti laajoille alueille ja saadessa paljon L2-puhujia. Tällaiset tutkimukset kumosivat oikeastaan kaksikin strukturalistisen lingvistiikan aksioomaa: paitsi sen, että kaikki kielet ovat yhtä monimutkaisia, myös sen, ettei kielen rakenne riipu yhteiskunnallisesta ympäristöstä. Kielten leviäminen ja niistä johtuvat kielikontaktit ja L2-puhujien lisääntyminen ovat kaikki kielten sosiologiaan liittyviä tapahtumia.

On kuitenkin tärkeää huomata, että kielikontaktien on todettu myös monimutkaistavan kieltä. Nichols (1992, 192–195) huomauttaa, että monilla maapallon alueilla, joissa kielellinen diversiteetti on korkea ja puhujat ovat usein monikielisiä, kielet ovat rakenteeltaan hyvin monimutkaisia. Esimerkkejä ovat Kaukasus, luoteinen Pohjois-Amerikka ja

Uusi-Guinea (vrt. Trudgill 2011, 29–32). Tällaisilla alueilla on paljon kielikontakteja, mutta mikään kieli ei ole levinnyt laajemman alueen lingua francaksi.

Trudgill (2011, 40–43) pyrkii selittämään, miksi kielikontaktit näyttävät toisinaan yksinkertaistavan, toisinaan monimutkaistavan kontaktitilanteisiin osallistuvia kieliä. Verrattain pinnallinen kontakti, johon liittyy ensi sijassa vain sanaston lainaamista, ei todennäköisesti muuta siihen osallistuvien kielten tyyppiä ollenkaan. Laajamittaisempi kontakti sen sijaan johtaa erilaisiin typologisiin muutoksiin sen mukaan, millaiseen yksilötason kaksikielisyyteen se perustuu. Samalla alueella puhuttavien kielten välinen pitkäaikainen kontakti, jossa lapsetkin ovat sujuvasti kaksi- tai monikielisiä, johtaa kielten monimutkaisuuden kasvuun, kun puhujat lisäävät kieliin toisista kielistä lainattuja piirteitä. Sen sijaan kontaktitilanne, jossa aikuiset joutuvat nopeasti opettelemaan heille oudon kielen, johtaa tämän kielen vähittäiseen yksinkertaistumiseen, jos L2-puhujia on kielen puheyhteisössä paljon tai jos heillä on korkea prestiisi (jos he ovat esimerkiksi vierasperäinen eliitti).

Balkanin kieliliittoa tutkinut Lindstedt (2019) on lisännyt tähän vielä kolmannen tyypin: Balkanin kielten pitkäaikainen kontakti keskiajan lopulla ja uuden ajan ensimmäisinä vuosisatoina perustui ennen kaikkea aikuisten stabiiliin pitkäaikaiseen monikielisyyteen. Sen tuloksia Balkanin kielissä ei voi pitää yksiselitteisesti yksinkertaistavina tai monimutkaistavina, vaan tuloksena oli ennen kaikkea kieliopillisten piirteitten eksplisiittinen merkintä analyyttisesti eli erilaisilla partikkeleilla, artikkeleilla ja prepositioilla. Kielet lainasivat toisiltaan uusia kieliopillisia piirteitä, esimerkiksi evidentiaalisuuden tai määräisyyden pakollisen merkinnän, mikä lisäsi kompleksisuutta sekä sääntöjen että redundanssin muodossa, mutta niiden merkinnässä lisääntyi läpinäkyvyys, mikä puolestaan oli yksinkertaistava muutos.

Kielikontaktien kielioppia yksinkertaistava vaikutus on ollut yksi selitysmalli monen eurooppalaisen kielen historian tutkimuksessa. Useimmat romaaniset kielet ovat kokonaan kadottaneet latinan sijataivutuksen, mikä voidaan selittää sillä, että latina levisi läntiseen ja eteläiseen Eurooppaan ensi sijassa kielenvaihdon kautta ja oli siis aikoinaan suurelle osalle puhujistaan toinen kieli (Clackson & Horrocks 2007, 276). Latinan

sijamuodot ovat osittain säilyneet romaniassa ja sen Balkanilla puhuttavissa lähisukukielissä, mikä taas voidaan selittää Balkanin muitten kielten sijajärjestelmän niille antamalla tuella verrattain stabiilissa kielikontaktitilanteessa, jossa ei tapahtunut pelkästään yksisuuntaista laajamittaista kielenvaihtoa (Lindstedt 2014, 172–175; Wahlström 2015, 130–131).

Englannin kielen historia on hyvä esimerkki siitä, että kontaktien vaikutusta kieliopin yksinkertaistumiseen ei ole helppo tutkia. Britanniassa puhuttiin kelttiläisiä kieliä englannin puhujien saapuessa 400–500-luvuilla. Sen jälkeen englantiin ovat vaikuttaneet ainakin viikinkien muinaisskandinaavi 700-luvulta lähtien, maan valloittaneiden normannien ranskan kieli vuodesta 1066 lähtien sekä vielä myöhempien maahanmuuttajien hollanti ja alasaksa (Thomason & Kaufman 1988, 265–269, 321–325; Filppula 2013). Jo klassisessa kielikontaktiteoksessaan Thomason ja Kaufman (1988, 263–331) käyvät laajasti läpi todistusaineistoa ja vertaavat englannin kehitystä muitten germaanisten kielten kehitykseen. Heidän johtopäätöksensä on, että englannin kieliopin yksinkertaistumista kantagermaanista lähtien ei voi pitää tuloksena kontaktista muinaisskandinaavin, ranskan tai myöhempien kontaktikielten kanssa vaan normaalina sisäsyntyisenä kehityksenä. McWhorter (2002) argumentoi kuitenkin, että nimenomaan skandinaavinen kontakti muutti voimakkaasti englannin kielioppia. Viime aikoina on puolestaan esitetty argumentteja sen puolesta, että englantia yksinkertaistikin varhaisempi kelttiläisten kielten ja englannin välinen kontakti ja kielenvaihto, joka jatkui Pohjois-Englannissa pitempään kuin aikaisemmin on ajateltu (Trudgill 2011, 50–55; Filppula 2013). McWhorter (2011, 261–295) on hänkin tarkastellut kelttiläistä vaikutusta, jota hän pitää varmana, vaikkei se hänen mukaansa ole ollut niin yksiselitteisesti kieltä yksinkertaistavaa kuin skandinaavinen vaikutus; hänen mukaansa englanti on semikreoli (ks. luku 3.5.).

Kielikontaktien typologinen vaikutus on siis ollut kiistanalaista jopa englannissa, jota on tutkittu paljon ja jonka sisäinen ja ulkoinen historia tunnetaan hyvin. Metodologinen opetus tästä on, että mitä laajemmista kieliopillisista muutoksista on kyse, sitä epävarmemmiksi kontaktipohjaiset selitykset käyvät, vaikkei olekaan mitään periaatteellista syytä, miksi kontaktiselitykset olisivat huonompia kuin kielen sisäisiin syihin

vetoavat selitykset. Luonnollisesti yhtä kieltä tutkittaessa ei voida olla oikopäätä varmoja siitä, mikä lopulta on ollut kielikontaktien ja kielen sisäisten, esimerkiksi sananloppuisten fonologisten muutosten keskinäinen osuus, mutta kun tietynlaisten sosiolingvististen kontaktitilanteiden osoitetaan johtavan eri kielissä säännöllisesti samanlaisiin muutoksiin (esimerkkinä edellä siteerattu Kusters 2003), kontaktiselityksen paino jokaisessa yksittäistapauksessakin kasvaa. Selitys voi olla myös sisäisten syiden ja useamman kuin yhden kontaktitilanteen yhteisvaikutus, kuten Wahlström (2015) on osoittanut tutkimuksessaan sijataivutuksen kadosta bulgariassa ja makedoniassa.

## 4. Yhteenveto: mitä kielikontaktien tutkijalta vaaditaan

Kuten tässä artikkelissa on käynyt ilmi, kielikontaktien tutkimus voi koskea mitä tahansa kielen tasoa, mikä johtuu puolestaan siitä, että kontaktilähtöinen kielenmuutos voi tapahtua millä tahansa kielen tasolla ääntämyksestä semantiikkaan ja pragmatiikkaan. Kielikontaktien tutkijan tulee näin ollen tuntea sitä kielitieteen alaa, joka hänen tarkastelemiaan ilmiöitä tutkii. Laajasti käytettyjä, helposti sovellettavia malleja ei kielikontaktien tutkimuksessa juurikaan ole, ja kontaktien tutkijan tarvitsemaa metodologista osaamista ei ole aivan yksinkertaista määritellä.

Sarah Thomason (2010) kuvaa artikkelissaan sitä, millaisen tutkimusprosessin kautta jonkin ilmiön kontaktilähtöisyys voidaan osoittaa. Prosessi ilmentää hyvin niitä erilaisia metodologisia valmiuksia, jotka voivat olla kontaktien tutkijalle tarpeen. Tutkittavia kieliä tulee tuntea kokonaisuutena, jotta voisi todeta, onko kielessä muita lainautuneita rakenteita. On myös kyettävä tarvittaessa osoittamaan, että ilmiön synnyn aikaan on vallinnut lainautumiselle otollinen sosiolingvistinen tilanne. On toisin sanoen tunnettava tutkittavan kielen kehityshistoriaa ja mahdollisesti myös puhujayhteisön kielenulkoista historiaa, tai jos kyse on tuoreesta ilmiöstä, sen nykyistä tilannetta.

Jotta taas kykenisi määrittelemään, mikä lainattua rakennetta lähtö- ja kohdekielessä yhdistää, on pystyttävä käyttämään laajemmin yleistettävissä olevia vertailukäsitteitä, jotka eivät nojaa pelkästään yhden kielen kuvaus- ja tutkimustraditioon. Tällaista otetta vaaditaan myös kontrastiivisessa kielentutkimuksessa, jossa kieliä vertaillaan myös ilman oletusta niiden välisestä kontaktivaikutuksesta (ks. kielten vertailun metodiikasta Kolehmainen, Miestamo & Nordlund 2013). Osoittaakseen, että kontakti-ilmiö todella on uusi kohdekielessä – tai vanha lähtökielessä –, tulee tutkijan verrata tutkittavia kieliä niiden sukukieliin tai varhaisempiin kielimuotoihin. Tällöin kyseeseen tulevat esimerkiksi historiallis-vertailevan kielentutkimuksen menetelmät, tai jos epäilty kontaktimuutos on nähtävissä kirjallisessa aineistossa, myös historiallinen korpustutkimus voi olla tarpeen (ks. Historiallinen korpuslingvistiikka tk.).

Edellä kuvatut menetelmälliset valmiudet eivät koske pelkästään kontaktitutkimusta. Usein kielikontaktien tutkijan erikoisosaaminen on tutkittavien kielten, niiden puhujayhteisöiden ja erityyppisten kielikontaktitilanteiden tuntemusta. Ei riitä pelkästään, että tuntee kielen, jossa tapahtuvia ilmiöitä havainnoi, vaan myös muutoksen lähteenä olevaa kieltä on syytä tuntea. Jos tarkastellaan synkronisia ilmiöitä ja käytetään esimerkiksi keskusteluaineistoa, korostuu tutkijan kielitaidon merkitys entisestään. Jos taas kyse on diakronisesta tutkimuksesta, on tunnettava myös niitä kielen varieteetteja, joihin otaksuttu kontakti ei ole vaikuttanut ja joiden voi näin ollen olettaa edustavan paremmin kontaktia edeltänyttä tilannetta. Toisaalta kun aineistona on typologinen piirreotos kuten makroalueiden tutkimuksessa, ei yksittäiskielten osaaminen ole välttämätöntä. Tällöin tarvitaan kuitenkin ymmärrystä kielityypeistä ja siitä, millaisin vertailukäsittein yksittäiskielten ilmiöt voidaan rinnastaa toisiinsa (ks. Kielitypologian menetelmät tk.).

Tässä artikkelissa on osoitettu, että kielikontakti-ilmiöiden selittäminen kietoutuu aina puhujayhteisöissä vallitsevien sosiolingvististen asetelmien ymmärtämiseen. Usein on yhtä tärkeää havainnoida kielenkäyttötilanteita kuin niitä kielen rakenteita, joissa ilmiö esiintyy. Joskus kontakti-ilmiön todistaminen jopa vaatii, että pystyy osoittamaan kontaktitilanteen olleen sen kaltaisen, että lainautuminen on ylipäätään ollut mahdollista. Myös monikielisten käytänteiden tarkastelu sinänsä

on osa kielikontaktitutkimusta laajassa mielessä. Kontaktiin johtaneen tilanteen ymmärtämisessä mikään ei voita sitä kokemusta, joka saavutetaan oleskelemalla tutkittavan puhujayhteisön parissa – silloinkin kun aineistoa ei hankita kenttätyössä.

Tutkija operoi aina myös intuitiollaan, jossa on kyse tiedon ja kokemuksen yhdistelmästä. Tuottaakseen tutkittavia hypoteeseja kontaktitutkijan tulee tuntea kontakti-ilmiöitä yleensä: mikä on mahdollista ja millaisissa olosuhteissa sekä millaisia syy-seuraus- ja muita riippuvuussuhteita ilmiöiden välillä voidaan havaita. Voidaankin sanoa, että keskeisen osan kontaktitutkijan metodologisesta osaamisesta muodostaa tutkimuskirjallisuuteen perehtyminen laajasti ja ennakkoluulottomasti, ja aloittaa voi esimerkiksi tämän artikkelin kirjallisuusluettelosta.

**Aiheesta lisää:**


Hickey, Raymond (toim.). 2010. *The Handbook of Language Contact*. Chichester, West Sussex & Malden (MA): Wiley-Blackwell.

Hickey, Raymond (toim.). 2017. *The Cambridge Handbook of Areal Linguistics*. Cambridge: Cambridge University Press.

*Journal of Language Contact*. Saatavissa: https://brill.com/view/journals/jlc/jlc-overview.xml.

*Journal of Pidgin and Creole Languages*. Saatavissa: https://www.jbe-platform.com/content/journals/15699870.

Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago and London: The University of Chicago Press.

Thomason, Sarah G. 2001. *Language Contact: An Introduction*. Washington (DC): Georgetown University Press.

Trudgill, Peter. 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press.

## VIITTEET

1 https://www.bbc.com/pidgin

## KIRJALLISUUS


Adamou, Evangelia. 2012. Social networks in Greek Thrace: Language shift and language maintenance. Julkaisussa: Lindstedt, Jouko & Wahlström, Max (toim.) *Balkan Encounters: Old and New Identities in South-Eastern Europe*. Slavica Helsingiensia 41. Helsinki: Helsingin yliopisto, 7–32. Saatavissa: https://blogs.helsinki.fi/slavica-helsingiensia/slavica-helsingiensia-41/.

Anttila, Raimo. 1972. *An Introduction to Historical and Comparative Linguistics.* New York (NY): Macmillan.

Arends, Jacques, Muysken, Pieter & Smith, Norval (toim.). 1995. *Pidgins and Creoles: An Introduction*. Creole Language Library 15. Amsterdam: John Benjamins.

Auwera, Johan van der. 1998. Conclusion. Julkaisussa: Auwera, Johan van der (toim.) *Adverbial Constructions in the Languages of Europe*. Empirical Approaches to Language Typology, EUROTYP 20-3. Berlin: Mouton de Gruyter, 813–836.

Bartens, Angela. 2009. Kreolikielistä. Julkaisussa: Kalliokoski, Jyrki, Kotilainen, Lari & Pahta, Päivi (toim.) *Kielet kohtaavat*. Tietolipas 227. Helsinki: Suomalaisen Kirjallisuuden Seura, 50–74.

Bloomfield, Leonard. 1933. *Language*. New York (NY): Henry Holt.

Boas, Franz. 1911. *Handbook of American Indian Languages*, osa 1. Washington: Smithsonian Institution, Bureau of American Ethnology.

Boduen-de-Kurtenè, Ivan A. 1901. O směšannom harakterě vsěh jazykov. Žurnal Ministerstva narodnago prosvěščenija 337, 12–24.

Broch, Ingvild & Jahr, Ernst Håkon. 1984. *Russenorsk: et pidginspråk i Norge.* (2. laitos.) Oslo: Novus.

Campbell, Lyle. 2006. Areal linguistics: A closer scrutiny. Julkaisussa: Matras, Yaron, MacMahon, April & Vincent, Nigel (toim.) *Linguistic Areas: Convergence in Historical and Typological Perspective*. New York (NY): Palgrave Macmillan, 1–31.

Clackson, James & Horrocks, Geoffrey. 2007. *The Blackwell History of the Latin Language*. Malden (MA): Blackwell.

Contact Language Library. John Benjamins. (Aiempi nimi: Creole Language Library.)

Dahl, Östen. 2004. *The Growth and Maintenance of Linguistic Complexity.* Studies in Language Companion Series 71. Amsterdam: John Benjamins.

Dahl, Östen. 2008. Kuinka eksoottinen kieli suomi on. *Virittäjä* 112:4, 545–559.

Darwin, Charles. [1874]. *The Descent of Man and Selection in Relation to Sex.* (2. laitos). [verkkoaineisto]. [viitattu 29.3.2018.] Saatavissa: http://www.gutenberg.org/ebooks/2300.

Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13:2, 257–292.

Edwards, John R. 1994. *Multilingualism*. London: Routledge.

Emeneau, Murray Barnson. 1956. India as a Lingustic Area. *Language* 32:1, 3–16.
Enfield, Nick J. 2005. Areal linguistics and mainland Southeast Asia. *Annual Review of Anthropology* 34, 181–206.
EUROTYP = Eurotyp 1 – Eurotyp 8 = *Empirical Approaches to Language Typology* 20-1 – 20-8. Mouton de Gruyter 1997–2011.
Feuillet, Jack. 2012. *Linguistique comparée des langues balkaniques*. Collection de manuels publiée par l'Institute d'études slaves. Paris: Institute d'études slaves.
Filppula, Markku. 2013. Contact and the early history of English. Julkaisussa: Hickey, Raymond (toim.) *The Handbook of Language Contact*. Chichester: Wiley-Blackwell, 432–453.
Fishman, Joshua A. 1964. Language maintenance and language shift as a field of inquiry: A definition of the field and suggestions for its further development. *Linguistics* 2:9, 32–70.
Haan, Ferdinand de. 2013. Semantic distinctions of evidentiality. Julkaisussa: Dryer, Matthew S. & Haspelmath, Martin (toim.) *The World Atlas of Language Structures Online*. [verkkoaineisto]. [viitattu: 4.12.2017]. Leipzig: Max Planck Institute for Evolutionary Anthropology. Saatavissa: https://wals.info/chapter/77.
Haarmann, Harald. 1970. *Die indirekte Erlebnisform als grammatische Kategorie. Eine eurasische Isoglosse*. Veröffentlichungen der Societas Uralo-Altaica, 2. Wiesbaden: Otto Harrassowitz.
Hakulinen, Lauri. 1968. *Suomen kielen rakenne ja kehitys*. (3. p.). Helsinki: Otava.
Hall, Robert A. 1946. Bartoli's 'neolinguistica'. *Language* 22:4, 273–283.
Haspelmath, Martin. 1998. How young is Standard Average European? *Language Sciences* 20:3, 271–287.
Haspelmath, Martin. 2001. The European linguistic area: Standard Average European. Julkaisussa: Haspelmath, Martin, König, Ekkehard, Oesterreicher, Wulf & Raible, Wolfgang (toim.) *Language Typology and Language Universals – Sprachtypologie und sprachliche Universalien – La typologie des langes et les universaux linguistiques*. Handbücher zur Sprach- und Kommunikationwissenschaft 20. Berlin: de Gruyter, 1492–1510.
Haspelmath, Martin & Tadmor, Uri (toim.). 2009. *Loanwords in the World's Languages: A Comparative Handbook*. Berlin: De Gruyter Mouton.
Heine, Bernd & Kuteva, Tania. 2006. *The Changing Languages of Europe*. Oxford: Oxford University Press.
Häkkinen, Kaisa. 1990. *Mistä sanat tulevat: Suomalaista etymologiaa*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Joseph, Brian D. 2010. Language contact in the Balkans. Julkaisussa: Hickey, Raymond (toim.) *The Handbook of Language Contact*. Malden (MA): Wiley-Blackwell, 618–633.
*Journal of Language Contact*. Saatavissa: https://brill.com/view/journals/jlc/jlc-overview.xml.
*Journal of Pidgin and Creole Languages*. Saatavissa: https://www.jbe-platform.com/content/journals/15699870.
Kallio, Petri. 2006a. On the earliest Slavic loanwords in Finnic. Julkaisussa: Nuorluoto, Juhani (toim.) *The Slavicization of the Russian North: Mechanisms and Chronology*. Slavica Helsingiensia 27. Helsinki: Helsingin yliopisto, 154–166.
Kallio, Petri. 2006b. Suomen kantakielen absoluuttista kronologiaa. *Virittäjä* 110:1, 2–25.
Kallio, Petri. 2007. How Uralic is Stadin Slangi? Julkaisussa: Blokland, Rogier & Hasselblatt,

Cornelius (toim.) *Language and Identity in the Finno-Ugric World*. Studia Fenno-Ugrica Groningana 4. Maastricht: Shaker Publishing, 176–191.

Kolehmainen, Leena, Miestamo, Matti & Nordlund, Taru (toim.). 2013. *Kielten vertailun metodiikka*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kopitar, Jernej. 1829. Albanische, walakische u. bulgarische Sprache. *Jahrbücher der Literatur (Wien)* 46, 59–106.

Kouwenberg, Silvia & Singler, John Victor (toim.). 2008. *The Handbook of Pidgin and Creole Studies*. Chichester: Wiley-Blackwell.

Kuronen, Mikko & Leinonen, Kari. 2001. Fonetiska skillnader mellan finlandssvenska och rikssvenska. Julkaisussa: Jönsson, Linda, Adelswärd, Viveca, Cederberg, Ann, Pettersson, Per A. & Kelly, Caroline (toim.) *Svenskans beskrivning 24. Förhandlingar vid tjugofjärde sammankomsten för svenskans beskrivning, Linköping den 22–23 oktober 1999*. Linköping: Linköping University Electronic Press, 125–138.

Kusters, Wouter. 2003. *Linguistic Complexity: the Influence of Social Change on Verbal Inflection*. Utrecht: LOT. [verkkoaineisto]. [viitattu 29.3.2018.] Saatavissa: http://www.lotpublications.nl/Documents/077_fulltext.pdf.

Lehtonen, Heini. 2015. *Tyylitellen: Nuorten kielelliset resurssit ja kielen sosiaalinen indeksisyys monietnisessä Helsingissä*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-1333-7.

Lindstedt, Jouko. 2000. The perfect – aspectual, temporal and evidential. Julkaisussa: Östen Dahl (toim.) *Tense and Aspect in the Languages of Europe*. Empirical Approaches to Language Typology, Eurotyp 20-6. Berlin: Mouton de Gruyter, 365–383.

Lindstedt, Jouko. 2004. Imaginary roots. *Finnisch-Ugrische Forschungen* 58:1–3, 307–313.

Lindstedt, Jouko. 2009a. Balkanin kieliliitto. Julkaisussa: Kalliokoski, Jyrki, Kotilainen, Lari & Pahta, Päivi. 2009 (toim.) *Kielet kohtaavat*. Tietolipas 227. Helsinki: Suomalaisen Kirjallisuuden Seura, 75–95.

Lindstedt, Jouko. 2009b. Esperanto – an East European contact language? Julkaisussa: Voß, Christian & Nagórko, Alicja (toim.) *Die Europäizität der Slawia oder die Slawizität Europas: Ein Beitrag der kultur- und sprachrelativistischen Linguistik*. Studies on Language and Culture in Central and Eastern Europe 2. München: Otto Sagner, 125–134.

Lindstedt, Jouko. 2014. Balkan Slavic and Balkan Romance: from congruence to convergence. Julkaisussa: Besters-Dilger, Juliane, Dermarkar, Cynthia, Pfänder, Stefan & Rabus, Achim (toim.) *Congruence in Contact-Induced Language Change*. Berlin: De Gruyter, 168–183.

Lindstedt, Jouko. 2019. Diachronic regularities explaining the tendency towards explicit analytic marking in Balkan syntax. Julkaisussa: Krapova, Iliyana & Joseph, Brian (toim.) *Balkan Syntax and (Universal) Principles of Grammar*. Trends in Linguistics, Studies and Monographs 285. Berlin: De Gruyter Mouton, 70–84.

Matras, Yaron. 2002. *Romani: A Linguistic Introduction*. Cambridge: Cambridge University Press.

McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5:2–3, 125–166.

McWhorter, John. 2002. What happened to English? *Diachronica* 19, 217–272.

McWhorter, John H. 2011. *Linguistic Simplicity and Complexity: Why Do Languages Undress?*

Language Contact and Bilingualism 1. Boston: De Gruyter Mouton.
McWhorter, John H. 2018. *The Creole Debate*. Cambridge: Cambridge University Press.
Meijer, Guus & Muysken, Pieter. 1977. On the beginnings of pidgin and creole studies: Schuchardt and Hesseling. Julkaisussa: Valdmann, Albert (toim.) *Pidgin and Creole Linguistics*. Bloomington: Indiana University Press, 21–45.
Miestamo, Matti. 2008. Grammatical complexity in cross-linguistic perspective. Julkaisussa: Miestamo, Matti, Sinnemäki, Kaius & Karlsson, Fred (toim.) *Language Complexity: Typology, Contact, Change*. Studies in Language Companion Series 94. Amsterdam: John Benjamins, 23–41.
Miestamo, Matti. 2017. Linguistic diversity and complexity. *Lingue e linguaggio* 16:2, 227–253.
Mufwene, Salikoko S. 2008. *Language Evolution: Contact, Competition and Change*. London: Continuum.
Muysken, Pieter. 2000. *Bilingual Speech: A Typology of Code-Mixing*. Cambridge: Cambridge University Press.
Muysken, Pieter. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition* 16:4, 709–730.
Myers-Scotton, Carol. 2002. *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University Press.
Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago (IL): The University of Chicago Press.
Nicolaï, Robert. 2014. À propos de Schuchardt, du mélange des langues et du contact. *Journal of Language Contact* 7:2, 211–249.
Palosaari, Naomi & Campbell, Lyle. 2011. Structural aspects of language endangerment. Julkaisussa: Austin, Peter K. & Sallabank, Julia (toim.) *The Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press, 100–119
Pellegrini, Giovanni Battista. 1980. Substrata. Julkaisussa: Posner, Rebecca & Green, John (toim.) *Romance Comparative and Historical Linguistics*. Berlin: De Gruyter, 43–73.
Schmid, Monika S. 2011. *Language Attrition*. Cambridge: Cambridge University Press.
Steinmetz, Katy. 2014. The absolute final, last word of the year: Because. Because grammar. *Time* 8.1.2014. [verkkoaineisto]. [viitattu 4.12.2017.] Saatavissa: https://newsfeed.time.com/2014/01/08/the-absolute-final-last-word-of-the-year-because/.
Thomason, Sarah G. 2001. *Language Contact: An Introduction*. Washington (DC): Georgetown University Press.
Thomason, Sarah G. 2010. Contact explanations in linguistics. Julkaisussa: Hickey, Raymond (toim.) *The Handbook of Language Contact*. Malden (MA): Wiley-Blackwell, 31–47.
Thomason, Sarah Grey & Kaufman, Terrence. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
Trubetzkoy, Nikolai S. 1923. Vavilonskaja bašnja I smešenie jazykov. *Evrazijskij vremennik* 3, 107–124.
Trubetzkoy, Nikolai S. 1928. [proposition 16]. Julkaisussa: *Actes du premier Congrès international de linguistes*, 17–18.
Trudgill, Peter. 1986. *Dialects in Contact*. Oxford: Blackwell.

Trudgill, Peter. 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity.* Oxford: Oxford University Press.

Wahlström, Max. 2015. *The Loss of Case Inflection in Bulgarian and Macedonian*. Slavica Helsingiensia 47. Helsinki: University of Helsinki. Saatavissa: https://blogs.helsinki.fi/slavica-helsingiensia/slavica-helsingiensia-47/.

WALS = Dryer, Matthew S. & Haspelmath, Martin (toim.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. [verkkoaineisto]. [viitattu 29.3.2018.] Saatavissa: https://wals.info.

Weinreich, Uriel. 1970 [1953]. *Languages in Contact: Findings and Problems*. The Hague: Mouton.

Whorf, Benjamin. 2012. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. Toim. John B. Carroll, Stephen C. Levinson & Penny Lee. Cambridge (MA): MIT Press.

Wiik, Kalevi. 2002. *Eurooppalaisten juuret*. Jyväskylä: Atena Kustannus.

# Luku 13 Kielten dokumentointi ja kieliopin kuvaus

*Lotta Jalava*
https://orcid.org/0000-0002-6400-7055

*Erika Sandman*
https://orcid.org/0000-0002-2341-2927

**Mikä?**

Dokumentoiva kielitiede pyrkii aiemmin tutkimattomien tai vähän tutkittujen kielten monipuoliseen ja -muotoiseen tallentamiseen. Dokumentoiva kielitiede on noussut tarpeesta tallentaa maailman kielellistä perintöä ja tukea uhanalaisten kielten puhujia kieltensä säilyttämisessä. Keskeisessä roolissa on primaarinen aineisto, tyypillisesti kenttätyössä kerättävät puhetallenteet, joista kootaan aineisto pysyväissäilytystä ja myöhempää käyttöä varten. Dokumentaatioon liittyy usein myös kielen kieliopin kuvaaminen ja referenssikieliopin laatiminen dokumentoitavasta kielestä kerätyn aineiston pohjalta.

**Katso myös:**

Luku 3 Haastattelu aineistonkeruun metodina sosiolingvistiikassa

Luku 5 Lingvistinen etnografia toimintayhteisöissä

Luku 7 Laadullinen aineistopohjainen kielentutkimus

Luku 14 Kielitypologian menetelmät

Luku 16 Antropologinen lingvistiikka

## 1. Johdanto

Maailmassa puhutaan noin 7 000 kieltä, joista vain murto-osa on kuvattu kattavasti (Simons & Fennig 2018). Suurta osaa maailman kielistä ei ole koskaan nauhoitettu, eikä niitä ole tallennettu kirjalliseen muotoon. Kielet ja niiden puhujamäärät ovat jakautuneet hyvin epätasaisesti: vuosituhannen vaihteessa maailman kielistä 96:ta prosenttia puhui ensikielenään vain neljä prosenttia maapallon väestöstä[1] (Crystal 2000, 14). *The Ethnologuen* vuoden 2018 painoksen mukaan lähes puolet maailman väestöstä puhuu ensikielenään jotakin maailman kymmenestä suurimmasta kielestä: mandariinikiinaa, espanjaa, englantia, arabiaa jne. (Simons & Fennig 2018). Monet kielet ovat vaarassa kadota ennen kuin niitä on tallennettu ja kuvattu tulevia sukupolvia ja maailman kulttuuriperinnön säilyttämistä varten. On arvioitu, että jopa 90 prosenttia maailman kielistä uhkaa kadota tämän vuosisadan loppuun mennessä (Bradley & Bradley 2002, xi). Lingvistiikan keskiöön onkin viimeistään 2000-luvulla noussut vahvasti dokumentoiva kielitiede, jonka pyrkimyksenä on koota ja tallentaa edustavia ja eettisesti ja teknisesti kestäviä aineistoja maailman kielistä (esim. Himmelmann 1998; 2006).

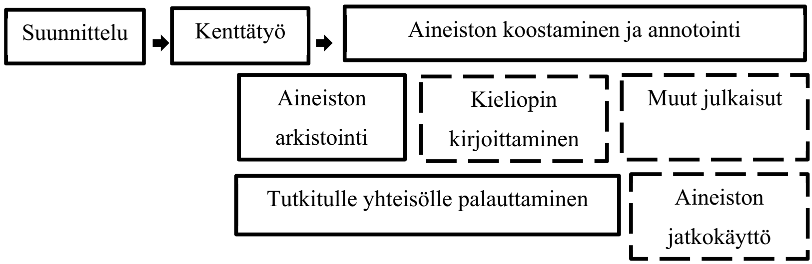
Dokumentoivan kielitieteen voidaan katsoa kuuluvan diversiteettilingvistiikkaan, jonka keskiössä on maailman kielellisen monimuotoisuuden kartoittaminen sekä kielten yhtäläisyyksien ja erojen tutkimus. Suuntaus yhdistää vähän tutkittujen kielten kuvaamista, kielten typologista vertailua sekä usein myös kielten historiallisten kehityskulkujen vertailua. Dokumentoivan kielitieteen perustana on kielitieteellinen kenttätyö, johon kuuluu aineiston kerääminen kielestä sen

luonnollisessa käyttöympäristössä vuorovaikutuksessa kielen puhujien kanssa, yhä useammin myös uhanalaisten kielten revitalisaation tarpeisiin (Austin & Sallabank 2018). Kielitieteellisen kenttätyön lähtökohdat ovat 1800-luvulta alkaen harjoitetussa antropologisessa lingvistiikassa (ks. Antropologinen lingvistiikka tässä kirjassa). Kenttätyön historiaan kytkeytyvät vahvasti amerikkalainen strukturalismi ja Pohjois-Amerikan lukuisten alkuperäiskielten kuvaaminen 1900-luvun alussa, toisaalta Suomessa taas fennougristinen kenttätyö ja materiaalien kerääminen suomen sukukielten kuvaamiseksi erityisesti 1800-luvun puolivälissä ja 1900-luvun alkuvuosina (ks. esim. Grünthal 2010; Saarinen 2007).

Dokumentoivan lingvistiikan tutkimusperinne rakentuu sellaisen kielikäsityksen varaan, jossa kieli nähdään osana puhujien henkistä, fyysistä, sosiaalista ja kulttuurista ympäristöä. Katoavien kielten ja kulttuurien dokumentointiin on kuulunut jo varhain sekä kielitieteellisen että etnografisen materiaalin keruu. Tutkimustiedon lisääntyessä on alettu ymmärtää yhä paremmin, millaiset monimutkaiset sosiokulttuuriset asetelmat määrittävät kielten ja niiden ympäristön suhdetta. Tarkasteltaessa kielen rakenteen, variaation ja muutoksen taustalla erottamattomasti vaikuttavia kulttuurisia, demografisia, taloudellisia ja poliittisia suhteita on puhuttu myös kielen ekologiasta (Haugen 1971). 2000-luvun kielentutkimuksessa on usein erotettu toisistaan **deskriptiivinen kielitiede** (engl. *descriptive linguistics*) ja **dokumentoiva kielitiede** (engl. *documentary linguistics*), joiden tutkimustehtävät ovat osittain erilaiset (Himmelmann 1998). Deskriptiivisessä kielitieteessä pääpaino on ennen kaikkea kieliopin kuvauksessa, kun taas dokumentoiva kielitiede keskittyy mahdollisimman laajojen ja monipuolisten kieliaineistojen keräämiseen ja julkaisemiseen tutkijoiden ja kieliyhteisön käyttöön. Kielen dokumentointiprojekti on parhaimmillaan usean tutkijan yhteistyö, joka tuottaa laajoja digitaalisia ääni- ja videokorpuksia sekä niiden tekstimuotoisia esityksiä käännöksineen ja analyyseineen, eikä aineistojen pohjalta välttämättä kirjoiteta kielioppia.

Tässä artikkelissa tarkastelumme kohteena on kenttätyö, jonka tavoitteena on jonkin aiemmin tutkimattoman tai vähän tutkitun kielen dokumentoiminen kieliopin kokonaiskuvauksen eli **referenssikieliopin** kirjoittamista ja kielennäytteiden kokoamista varten. Tarkastelemme

kielten dokumentointiin ja aineiston keräämiseen liittyvää metodologiaa sekä referenssikieliopin kirjoittamisen periaatteita ja käytänteitä. Uhanalaisten kielten dokumentointiin liittyy kiinteästi ajatus kielellisen monimuotoisuuden tallentamisesta ja säilyttämisestä, jolloin korostuu äidinkielisten puhujien osuus paitsi kenttätyöaineiston tuottamisessa myös sen jatkokäytössä kielen revitalisoinnin ja voimauttamisen näkökulmasta. Tutkijan kannalta taas kauemmas ja vieraaseen ympäristöön suuntautuva kenttätyömatka vaatii erityisen tarkkaa käytännön suunnittelua ja esivalmistelua. Kohteesta ja aihepiiristä riippumatta kaikenlainen kenttätyö sisältää aina suunnitteluvaiheen, varsinaisen aineistonkeruun sekä materiaalin koostamisen aineistoksi ja julkaisuiksi. Seuraavalla kuvalla havainnollistamme, että kenttätyöhankkeen ydin, itse kentällä tapahtuva työskentely, on vain yksi osa kielen dokumentointiin tai kenttätyöpohjaiseen tutkimukseen liittyvää hanketta.




Kaavio 1. Kielen dokumentaatiohankkeen vaiheet.

Antaaksemme lukijalle kuvan siitä, mihin kenttätyöllä tähdätään ja millaisia sisältöjä kenttätyöskentelyssä tutkitaan, kerromme tässä artikkelissa ensin yhdestä kielenkuvaushankkeen tyypillisestä tuloksesta, referenssikieliopista, ja sen tavoitteista, sisällöstä ja rakenteesta (luku 2). Kieliopin kuvaamisesta palaamme kentälle, kuvausta edeltävään aineistonkeruuvaiheeseen. Tutkimustarkoitusta vastaavan, monipuolisesti jatkokäytettävän ja tutkimuseettisesti kestävän kenttäaineiston kokoaminen vaatii huolellista työskentelyä paitsi kielenpuhujien parissa myös ennen kenttämatkalle lähtemistä sekä varsinaisen kenttätyön jälkeen. Luvussa 3 pohdimme kenttätyön etukäteissuunnittelua, ja luvussa 4

pääsemme varsinaisten kenttätyössä käytettävien aineistonkeruumetodien pariin. Luvussa 5 käsittelemme kenttämateriaalin työstämistä aineistoksi monipuolisen jatkokäytön näkökulmasta. Luku 6 on lyhyt yhteenveto.

# 2. Kenttätyöhön perustuva kieliopin kuvaaminen

Kielitieteellisen kenttätyön päämääränä on usein kokonaiskuvauksen laatiminen tutkittavan kielen kieliopista eli **referenssikielioppi**. Se on kattava yleiskuvaus kuvattavan kielen kaikista keskeisistä osa-alueista, kieliopillisista rakenteista ja niiden kommunikatiivisista funktioista esimerkkilausein havainnollistettuna. Lisäksi kieliopin kuvaukseen sisältyy tietoa puhujien fyysisestä ja sosiokulttuurisesta ympäristöstä ja näiden suhteesta tutkittavaan kieleen sekä kielennäytteitä, kuten otteita tallennetuista teksteistä. Referenssikieliopin kirjoittamista voidaan monessa mielessä pitää vähän kuvatun kielen tutkimuksen perustana. Kokonaiskuvaus tutkittavasta kielestä antaa hyvän pohjan yksittäisten kieliopin ilmiöiden syvälliselle tarkastelulle. Riittävän laaja, laadukas ja monipuolinen kielioppi voi toimia myös kielenhuollon tukena yhteiskunnan tarpeita varten. Tässä luvussa tarkastelemme kuitenkin ennen kaikkea kieliopin kirjoittamista tutkimuskäyttöä ja kielen ominaisuuksien tallentamista ajatellen.

## 2.1. Referenssikielioppi – mikä ja miksi?

Kielioppeja voidaan kirjoittaa hyvin monia erilaisia käyttötarkoituksia varten, esimerkiksi opetus- ja opiskelu- tai tutkimuskäyttöön, kielen normittamista ajatellen tai kielen ominaisuuksien tallentamiseksi osana maailman kulttuuriperintöä. Käyttötarkoituksesta ja kohdeyleisöstä riippuu, kuinka laaja kieliopista tehdään ja miten sen sisältö jäsennellään.

Referenssikielioppi on luonteeltaan ennen kaikkea hakuteos, josta lukija voi halutessaan löytää tietoa jostakin tutkittavan kielen kieliopillisesta ilmiöstä, kuten demonstratiiveista, aspektista tai passiivilauseista. Referenssikieliopeista ovat kiinnostuneita monenlaisista eri taustoista tulevat tutkijat, esimerkiksi eri kielikuntien asiantuntijat, kielitypologit, historiallisen kielitieteen, kielikontaktien ja areaalilingvistiikan tutkijat sekä kielioppiteorioiden laatijat mutta myös historioitsijat, kulttuurintutkijat ja yhteiskuntatieteilijät. Joissakin tapauksissa myös itse kieliyhteisö voi hyödyntää referenssikielioppia esimerkiksi kielenhuollon, kirjoitusjärjestelmän kehittämisen tai koulutuksen suunnittelun apuna. Referenssikielioppien olisi siis sovittava laajalle lukijakunnalle. Lisäksi referenssikielioppeja käytetään usein hyvinkin pitkien aikojen kuluttua niiden laatimisesta. Jos tutkittava kieli on hyvin uhanalainen, referenssikielioppi saattaa jäädä ainoaksi lähteeksi, joka siitä on tulevaisuudessa käytettävissä.

Ihannetapauksessa kielioppi palvelee sekä tutkijoita että kieliyhteisöä, mutta molempien kohdeyleisöjen puhutteleminen samassa teoksessa voi olla haastavaa jo esimerkiksi terminologisten ja ortografisten valintojen vuoksi. Ratkaisuna voisi olla, että kieliopin kirjoittaja laatii sekä tutkijayhteisölle suunnatun kieliopin että kieliyhteisölle suunnattua materiaalia (kielioppi, tekstikokoelmia, sanakirja). Hyvä tutkimusetiikka edellyttää, että tutkimuksessa käytetyt aineistot palautetaan kieliyhteisölle sellaisessa muodossa, että yhteisö voi niitä hyödyntää. Referenssikieliopin kirjoittamisen lisäksi kielen dokumentointiin kuuluu erottamattomasti sanaston kerääminen sekä kielennäytteiden julkaiseminen teksteinä ja nauhoitteina. Referenssikielioppi, sanakirja ja tekstikokoelma ovat kielenkuvauksen kolme kulmakiveä; näitä nimitetään usein *boasilaiseksi trilogiaksi* amerikkalaisen strukturalistin Franz Boasin mukaan, joka omassa työssään pyrki edistämään kaikkien näiden kolmen osa-alueen ottamista mukaan kielenkuvaukseen (Evans & Dench 2006, 10). Myös nykyään kielten dokumentointiin tulisi ihannetapauksessa kuulua niin kieliopin kuvausta, sanastotyötä kuin kieliaineistojen julkaisemistakin. Laajemman dokumentaatioprojektin yhteydessä voidaan tuottaa erilaisia digitaalisia ääni-, video- ja tekstimateriaaleja ja kielenainesten analyysia yhdistäviä verkkoaineistoja ja -sanakirjoja.

Kuitenkaan kaikissa hankkeissa ei käytännön syistä ole mahdollista keskittyä kaikkiin kielten dokumentoinnin osa-alueisiin yhtä syvällisesti.

Koska referenssikielioppien lukijakunta on laaja ja kielioppeja käytetään usein pitkienkin aikojen kuluttua niiden laatimisesta, niiden kirjoittamisessa olisi mahdollisuuksien mukaan pyrittävä esitystapaan, joka on helposti lähestyttävä monenlaisista taustoista tuleville tutkijoille. Tämän vuoksi referenssikielioppien laatimisessa ei nykyään yleensä käytetä kovin abstrakteja teoreettisia kielioppimalleja, jotka ovat ymmärrettäviä vain asiaan perehtyneille tutkijoille ja joihin perustuvat esitykset vanhenevat helposti teorioiden kehittyessä. Sen sijaan niiden laatimisessa käytetään verrattain epämuodollista, deskriptiivistä teoriakehystä, jota kutsutaan *peruskielioppiteoriaksi* (engl. *Basic Linguistic Theory*, BLT) (Dixon 2010; Dryer 2006; ks. myös Antropologinen lingvistiikka tk.). Peruskielioppiteoria on saanut paljon vaikutteita kielitypologiasta, ja se sisältää useimpien kielioppiteorioiden jakamia, osin myös koulujen kieltenopetuksen perusteella ymmärrettäviä kieliopillisia käsitteitä, kuten *subjekti*, *adjektiivi*, *postpositio* ja *sijamuoto*.

Kieliopin kirjoittajalla on ohjenuoranaan aina aineistolähtöisyys ja kielen kuvaaminen mahdollisimman luonnollisessa käyttökontekstissaan. Aineistona suositaan tekstejä ja keskusteluja, ja niistä esiin nousevien ilmiöiden perusteella tehdään yleistyksiä kieliopillisesta järjestelmästä. Kieliopin kuvaamisessa tehdyt ratkaisut on aina perusteltava suhteessa käytettyyn aineistoon; jokaisesta kieliopissa kuvatusta ilmiöstä on annettava runsaasti aineistoesimerkkejä.

Referenssikielioppien laatimisella on erityisen läheinen suhde kielitypologiaan. Kielitypologit käyttävät aineistonaan enimmäkseen referenssikielioppeja, ja toisaalta taas kielitypologien tuottama tieto siitä, mikä maailman kielissä on mahdollista, tavallista tai epätavallista, auttaa yksittäiskielen tutkijaa kieliopin laatimisessa. Referenssikieliopeissa kuvataan usein sellaisia kieliä, joilla ei ole vahvaa tutkimustraditiota ja vakiintunutta kielioppiterminologiaa. Kuvausmallin laatimisessa hyödynnetään silloin kielitypologien tekemiä yleistyksiä maailman kielistä. Tällaista tietoa voisi olla esimerkiksi se, mitä sanaluokkia maailman kielistä on löydetty ja miten eri sanaluokat tyypillisesti taipuvat. Useimmissa kielissä tehdään esimerkiksi ero substantiivien ja verbien välillä,

joten on todennäköistä, että tämä ero tehdään jollakin tavalla myös tutkittavassa kielessä. Kielitypologisen tutkimuksen pohjalta on laadittu erilaisia oppaita ja "tarkistuslistoja" kenttätyön ja kieliopin kirjoittamisen avuksi (Comrie & Smith 1977; Shopen 1985; 2007; Payne 1997). On kuitenkin tärkeää huomioida, että vaikka typologinen tieto on kieliopin kuvauksessa tärkeää, referenssikieliopin kirjoittamisen päämääränä on aina tutkittavan kielen kuvaaminen sen omista lähtökohdista käsin. Maailman kieliä koskevien yleistysten, teorioiden tai kielioppien kirjoitusoppaiden ei pidä antaa ohjata työskentelyä liikaa. Esimerkiksi englannin kieliopin kuvauksessa kieliopillinen subjekti on erittäin tärkeä kategoria, kun taas mandariinikiinassa suuri osa lauserakenteista on luonteeltaan topiikki-kommenttirakenteita ja subjekti ei ole yhtä tärkeä kieliopillinen kategoria kuin monissa muissa kielissä. Mandariinikiinan kieliopin kirjoitustyössä painotetaan tällöin topiikkirakenteiden kuvausta, eikä lauseissa oleteta automaattisesti olevan subjekteja ainoastaan sillä perusteella, että subjekti on maailman kielissä yleinen kieliopillinen kategoria (ks. Li & Thompson 1981).

Kieliopin kuvauksessa keskitytään niihin rakenteisiin ja ilmiöihin, jotka ovat oleellisia juuri tutkittavassa kielessä. Typologinen tieto auttaa yksittäiskielen tutkijaa ennen kaikkea kysymyksenasettelussa sekä kenttätyön ja kieliopin rakenteen suunnittelussa (Sandman 2013, 56). Lisäksi typologisen tiedon avulla tutkittavan kielen ilmiöt voidaan asettaa laajempaan viitekehykseen; mikä tutkittavassa kielessä on erityisen poikkeuksellista ja kiinnostavaa kieltenvälisestä näkökulmasta. Esimerkiksi monissa tiibetiläiskielissä on mahdollista ilmaista pakollisen verbimorfologian avulla, onko puhuja itse osallistunut tapahtumaan, josta kertoo, vai tarkasteleeko hän tapahtumaa ulkopuolisen roolissa (ks. San Roque, Floyd & Norcliffe 2018, 2). Tällainen ns. egoforisuus- tai perspektiivikategoria on löydetty vain hyvin harvoista maailman kielistä, ja siksi se on typologisesta näkökulmasta erityisen mielenkiintoinen ilmiö, jota kannattaa tuoda tiibetiläiskielten referenssikieliopeissa esiin (ks. myös luku 2.2.). Toisaalta taas kielitypologien pääasiallista aineistoa ovat referenssikieliopit; koska typologisen tutkimuksen aineistot ovat laajoja ja käsittävät yleensä useita satoja kieliä, typologi ei voi osata tutkimiaan kieliä itse vaan on riippuvainen muiden tutkijoiden laatimista

kielenkuvauksista (ks. Kielitypologian menetelmät tk.). Referenssikieliopit olisikin pyrittävä laatimaan niin, että niitä aineistona käyttävä typologi löytää helposti etsimänsä tiedon (ks. luku 2.3.). Hedelmällinen yhteistyö yksittäisten kielten kieliopin kirjoittajien ja kielitypologien välillä hyödyttää molempia osapuolia.

Sen lisäksi että kielioppien lukijakunta on laaja, on myös kielioppien kirjoittajissa monenlaisista taustoista tulevia tutkijoita opiskelijoista urallaan pitkälle ehtineisiin asiantuntijoihin. Referenssikieliopin laatiminen vähän kuvatusta kielestä on aina pitkäaikainen ja vaativa projekti, jossa tarvitaan monenlaisia tutkimustaitoja sekä lisäksi tietynlaista pitkäjänteistä omistautumista tutkimustyötä kohtaan. Materiaalin kerääminen kentällä, kenttämateriaalien työstäminen tutkimusaineistoiksi ja kieliopin kirjoittaminen on työlästä ja aikaa vievää. Lisäksi tutkimuksen toteutukseen vaikuttavat lukuisat kenttätyöolosuhteisiin liittyvät tekijät, joihin tutkija ei aina pysty itse vaikuttamaan. Kokonaisen vähän tutkitun kielen referenssikieliopin kirjoittaminen yksin väitöskirjatyönä on mahdollista, ja tällaisia väitöskirjoja on tehty jonkin verran mm. Helsingin yliopistossa (ks. Wilde 2008; Berghäll 2010; Sippola 2012; Killian 2015; Sandman 2016; Honkasalo 2019; Yliniemi 2019). Myös maisterintutkinnon puitteissa on mahdollista aloittaa kieliopin kirjoittaminen, tutkia jotakin kieliopin osa-aluetta kenttätyön avulla tai vaikkapa laatia kielioppiluonnos jostakin vähän kuvatusta kielestä, mutta kenttätyöhön perustuva opinnäytetyö vaatii aina erityisen huolellista suunnittelua ja paljon tukea ja ohjausta. Olisi myös erittäin toivottavaa, että yksin puurtamisen sijaan kieliopin kuvausta ja kielten dokumentointia tehtäisiin tulevaisuudessa entistä enemmän tutkimusryhmissä. Ryhmissä työskentely mahdollistaa tehtävien jakamisen, laajempien tutkimusaineistojen kokoamisen ja neuvojen saamisen kokeneemmilta tutkijoilta ja takaa siten paremmin myös tutkimuksen validiteetin. Jos kieliopin kuvaamisessa suosittaisiin yksin työskentelyn sijaan enemmän ryhmätyöskentelyä, myös kenttätyöhön perustuvien opinnäytetöiden tekeminen olisi jatkossa helpompaa. Myös kenttätyömetodiikan opetuksen lisääminen yliopistoissa sekä opettajien ja opiskelijoiden yhteiset kenttäretket ovat hyvä tapa parantaa tulevien tutkijoiden valmiuksia kenttätyöhön perustuvaan kieliopin kuvaukseen. Parhaimmillaan kenttätyöhön perustuva

kielten dokumentointi ja kieliopin kuvaus on erittäin antoisaa ja palkitsevaa työtä, jonka tieteellinen ja yhteiskunnallinen merkitys on suuri.

## 2.2. Kieliopin rakenne ja osa-alueet

Koska eri kielten kieliopillisissa rakenteissa on huomattavia eroja, myös referenssikielioppien rakenteessa tehdyt ratkaisut vaihtelevat. Paitsi tutkittava kieli ja kieliopin kohdeyleisö myös tutkijan oma tausta ja henkilökohtaiset mieltymykset vaikuttavat kieliopin kirjoittamiseen. Kieliopin kirjoittaminen on aina ainutkertainen prosessi. Kieliopin kirjoittaja voi esimerkiksi joutua valikoimaan tai kehittelemään uutta terminologiaa harvinaisille ilmiöille, joille ei ole vakiintuneita nimityksiä typologisessa tutkimuksessa. Toisaalta referenssikielioppien jäsentelyssä voidaan hahmottaa myös yleisiä periaatteita, joita käsittelemme tässä artikkelissa. Tieto referenssikielioppien tyypillisestä rakenteesta auttaa niin aloittelevaa kieliopin kirjoittajaa työn suunnittelussa kuin kieliopin lukijaa kieliopin laadun arvioinnissa, mutta kieliopin kirjoittaja joutuu aina erikseen miettimään myös niiden ilmiöiden kuvausta, jotka eivät sovi olemassa oleviin malleihin.

On tärkeää pitää mielessä, että referenssikielioppi on aina kokonaisuus, jossa täytyy käsitellä kaikki kieliopin keskeiset osa-alueet. Kieliopin kirjoittaja ei voi sivuuttaa sellaisiakaan kielenkuvauksen osa-alueita, joiden kuvaaminen on erityisen vaikeaa tai joita koskeva tieto on riittämätöntä. Kielioppi, joka ei käsittele lainkaan vaikkapa intransitiivilauseita tai negaatiota, on epätäydellinen. Verrattain suppeakin referenssikielioppi on aina laaja teos, pituudeltaan 300–400 sivua tai enemmän. Jos laajamittainen kenttätyö ja kokonaisen kieliopin kirjoittaminen ei jostain syystä ole mahdollista, voidaan tutkittavasta kielestä laatia myös lyhempi kuvaus, kielioppiluonnos (engl. *grammar sketch*). Kielioppiluonnoksen pituus on yleensä joitakin kymmeniä sivuja. Sen rakenne on periaatteessa samankaltainen kuin referenssikieliopin, mutta se on paljon suppeampi esitys tutkittavasta kielestä ja antaa vain perustiedot ja joitakin esimerkkejä käsiteltävistä aihepiireistä[2]. Tällainen kielenkuvaus antaa tutkittavasta kielestä vain pintapuoliset tiedot mutta voi toimia hyvänä pohjana laajemmalle tutkimukselle.

Seuraavassa on lueteltu kielenkuvauksen keskeisimpiä osa-alueita, joihin referenssikieliopeissa usein otetaan kantaa. Nämä voivat joko olla omina lukuinaan kieliopissa tai niitä voidaan yhdistellä toisiinsa tai jakaa alalukuihin. Kaikissa kielissä on äänteitä ja monenlaisia muotokategorioita, sanat jakautuvat yleensä sanaluokkiin, sanoista voidaan muodostaa lauseita ja lauseita yhdistellään laajemmiksi kokonaisuuksiksi. Siksi näihin asioihin otetaan yleensä kieliopissa kantaa, jotta kielenkuvaus olisi kattava ja vertailukelpoinen muiden kielenkuvausten kanssa. Kuitenkin kielissä on suuria eroja sen välillä, kuinka paljon ja millaisia äänteitä esiintyy, miten muotokategorioita käytetään, mitä sanaluokkia siinä mahdollisesti on ja miten ne taipuvat ja millaisiin tyyppeihin lauseet voidaan jakaa. Näiden yksityiskohtien esittäminen eroaa myös eri kielten kieliopeissa suuresti, ja modernin funktionaalisen kielitieteen lähtökohtana on aina tuoda esiin juuri tutkittavalle kielelle ominaiset ilmiöt. Tästä syystä alla esitettyä tulee pitää vain suuntaa antavana mallina referenssikieliopeissa usein käsitellyistä asioista, ei varsinaisena kieliopin dispositioehdotuksena, joka olisi sellaisenaan sovellettavissa minkä tahansa kielen kuvaukseen. Kieliopeissa tehdyt ratkaisut nousevat aina tutkijan omassa aineistossaan havaitsemista säännönmukaisuuksista.

Referenssikieliopit aloitetaan yleensä **johdannolla**. Johdannossa kerrotaan perustiedot tutkittavasta kielestä ja tutkimuksen taustasta. Perustietoja ovat tutkittavan kielen maantieteellinen, genealoginen ja kielisosiologinen konteksti: Missä tutkittavaa kieltä puhutaan, mihin kielikuntaan se kuuluu ja millainen on sen puhuma-alueen kielellinen tilanne? Mitä kieliä tutkittavan kielen puhujat osaavat, miten he käyttävät eri kieliä arkielämässään ja millainen on tutkittavan kielen asema yhteisössä? Mitä nimityksiä tutkittavasta kielestä ja sen puhujista käytetään? Kokevatko kielen puhujat kuuluvansa johonkin etniseen ryhmään? Monissa maissa on käytössä myös jonkinlainen virallinen etnisten ryhmien luettelo (esim. Kiinassa on 55 virallisesti tunnustettua kansallista vähemmistöä), joka voi olla enemmän tai vähemmän keinotekoisesti määritelty eikä vastaa kieliyhteisöjen omaa identiteettiä. Johdantoluvussa esitellään lyhyesti myös alueen historiaan sekä kieliyhteisön kulttuuriin, kuten elinkeinoihin, tapakulttuuriin ja uskontoihin, liittyviä kysymyksiä. Tämä osuus referenssikielioppien johdannossa on pakostakin

yksinkertaistettu esitys aiheesta, koska tutkimuksen pääpaino on kieliopin kuvauksessa.

Kielioppien johdantoluvuissa esitetään yleensä myös lyhyt ja tiivis läpileikkaus tutkittavan kielen kieliopista ja sille ominaisista typologisista piirteistä: perussanajärjestys, morfologinen typologia, pääsanan ja määritteiden merkintä (ns. *head/dependent-marking*), kiinnostavimmat kieliopilliset kategoriat jne. Kielisosiologisen ja typologisen taustatiedon lisäksi johdantoon kuuluu tutkimuksen teoreettinen tausta, joka moderneissa referenssikieliopeissa on useimmiten peruskielioppiteoria. Kieliopin kirjoittaja kertoo lisäksi johdannossa tutkittavaa kieltä koskevasta aikaisemmasta tutkimuksesta, omasta kenttätyöstään ja sen taustasta: Millaista tutkimusta kyseisestä kielestä on aiemmin tehty ja miten referenssikielioppi suhteutuu siihen? Milloin ja kuinka pitkään kieliopin kirjoittaja on tehnyt kenttätyötä tutkittavan kielen puhujien parissa ja millaisia kielenoppaita hänellä on ollut apunaan? Millainen on kielenoppaiden asema yhteisössä? Kuinka laaja kieliopin pohjana käytetty aineisto on (esim. kuinka monta tuntia puhetta on nauhoitettu, kuinka monta transkriboitua lausetta kieliopin aineisto sisältää) ja millaista aineistoa on kerätty, mukaan lukien esim. kertomukset, keskusteluaineisto, elisitoidut lauseet ja stimulusten käyttöön perustuva aineisto (ks. myös luku 4)?

Useimmissa kieliopeissa käsitellään johdannon jälkeen äänneoppia eli **fonologiaa**, esimerkiksi millainen on tutkittavan kielen äännejärjestelmä, missä asemissa äänteet esiintyvät, millaista allofonista vaihtelua esiintyy, millainen on tavurakenne ja millainen rooli on prosodialla. Toinen tärkeä kielenkuvauksen osa-alue on **muotokeinot**: millaisia muotokeinoja tutkittava kieli käyttää (esim. affiksaatio, reduplikaatio, sisäinen muunnos) ja millaista morfofonologista vaihtelua esiintyy.

Oleellinen asia minkä tahansa kielen kieliopin kuvauksessa ovat **sanaluokat**: Mitä sanaluokkia kielessä on (tai onko sanaluokkien määrittely ylipäätään mahdollista) ja millaisin kriteerein sanaluokat voidaan erottaa toisistaan? Löytyykö eri sanaluokista niille tyypillisiä muotokategorioita? Sanaluokat ovat aina mitä suurimmassa määrin kielikohtaisia. Osassa maailman kieliä ei ole esimerkiksi adjektiiveja yhtenä erillisenä sanaluokkana. Esimerkiksi suomen etäsukukieliin kuuluvassa

tundranenetsissä ominaisuutta ilmaisevat sanat jakaantuvat kahteen sanaluokkaan ja taipuvat joko kuten nominit tai kuten verbit: esimerkiksi 'musta' on verbi ja 'valkoinen' nomini[3] (Jalava 2013). Toisaalta taas monissa Papuan ja Australian kielissä vaikkapa sukulaisuustermit ovat oma sanaluokkansa (esimerkki papualaisen oksapminin kielen sanaluokista ks. Loughnane 2010). Kieliopin kirjoittajan on pohdittava, millainen sanaluokkajako kielessä on ja kuinka monta käsittelylukua sanaluokkien kuvaamiseen tarvitaan. Kielitypologisen tutkimuksen valossa sanaluokista universaaleimmat ovat substantiivi ja verbi (Payne 1997, 33). Niiden välinen ero voidaan tehdä useimmissa tunnetuissa kielissä, ja siksi moniin referenssikielioppeihin sisältyvät luvut substantiivilausekkeen ja verbilausekkeen rakenteesta. On kuitenkin kieliä, joissa sanaluokkien erottaminen toisistaan saattaa olla hyvinkin ongelmallista ja edes substantiivin ja verbin välinen ero ei ole selkeä. Tällaisia kieliä ovat etenkin rakenteeltaan hyvin analyyttiset kielet, joissa on vain niukasti tai ei ollenkaan taivutusmorfologiaa. Ääriesimerkki hyvin analyyttisestä kielestä on riaunindonesia, jossa substantiiveille ja verbeille ominaisten muotokategorioiden erottaminen on lähes mahdotonta (ks. Gil 2001, 118–119).

Sananlaajuisten yksikköjen lisäksi kieliopeissa käsitellään tyypillisesti **lauseita**: millainen on tutkittavan kielen perussanajärjestys, millainen on väitelauseen rakenne, mitä muita lausetyyppejä kuin väitelauseita esiintyy (esim. kysymykset, käskyt) sekä millaisin keinoin lauseita yhdistetään? Lauseoppia ei aina ole mahdollista käsitellä yksityiskohtaisesti, mikäli tutkittava kieli on erittäin uhanalainen eivätkä puhujat käytä sitä aktiivisesti arkielämässään. Tällöin ainoaksi aineistonkeruumenetelmäksi saattaa jäädä yksittäisten sanojen ja hyvin yksinkertaisten lauseiden kysyminen suoraan puhujilta, ja aineiston suppeus näkyy myös valmiissa kieliopissa, josta jäävät tällöin puuttumaan ainakin monimutkaisemmat lauserakenteet.

Muodollisten kategorioiden, kuten sanojen ja lauseiden, lisäksi moderneissa referenssikieliopeissa käsitellään omina lukuinaan runsaasti myös monenlaisia **kieliopillisia merkityksiä ja funktioita**, kuten valenssin, aikamuotojen, aspektin, topiikkirakenteiden ym. ilmaisemista. Muodon ja merkityksen roolia kieliopin kuvauksessa käsitellään tarkemmin luvussa 2.3.

Kielioppeihin kuuluu lisäksi **lähdeluettelo**, **tekstinäytteitä eri tekstilajeja edustavista teksteistä** sekä mahdollisesti **hakemisto kieliopissa käytetyistä asiasanoista**. Viimeksi mainittua ei monissa referenssikielioppeissa ole, mutta hakemiston laatiminen auttaa lukijaa löytämään haluamansa tiedon.

## 2.3. Muotolähtöisyys sekä merkitys- ja funktiolähtöisyys kieliopin kuvauksessa

**Kaksi referenssikieliopin esimerkkidispositiota**

Dispositio 1 (muotolähtöinen, perustuu löyhästi Wilden [2008] kielioppiin).

1. Johdanto
2. Äänneoppi
3. Muoto-oppi
4. Substantiivilauseke
5. Verbilauseke
6. Adjektiivit
7. Adverbit
8. Muut sanaluokat
9. Lauserakenne
10. Yhdyslauseet
11. Tekstinäytteitä

Dispositio 2 (enemmän merkitys- ja funktiolähtöinen, perustuu löyhästi Eppsin [2008] kielioppiin).

1. Johdanto
2. Äänneoppi
3. Sanaluokat
4. Nominimorfologia
5. Verbimorfologia
6. Argumenttien merkintä

7. Aspekti
8. Modaalisuus ja evidentiaalisuus
9. Lauserakenne
10. Negaatio
11. Kysymykset
12. Yhdyslauseet
13. Tekstinäytteitä

Kielioppien sisällön jäsentely voi perustua toisaalta muotoon ja toisaalta merkitykseen ja kieliopilliseen funktioon. Muotokeinojen, sanaluokkien ja lauserakenteiden kuvaaminen omissa luvuissaan on esimerkki **muotolähtöisestä** lähestymistavasta, jolle on tyypillistä kuvauksen jäsentely esimerkiksi morfologian ja taivutusparadigmojen mukaan. Useimmissa kieliopeissa on ainakin joitakin lukuja, joissa lähtökohtana on muoto, esimerkiksi sija- tai persoonataivutus. Näiden lisäksi modernissa deskriptiivisessä kielitieteessä painotetaan kuitenkin vahvasti myös **merkitys- ja funktiolähtöistä** lähestymistapaa (ks. Ojutkangas, Larjavaara, Miestamo & Ylikoski 2009), jossa lähtökohtana on muodon sijaan jokin kielellinen funktio (esim. modaalisuus, aspekti, informaatiorakenne, deiksis jne.). Sen alla käsitellään erilaisia kieliopillisia rakenteita, joilla voidaan ilmaista kyseistä funktiota. Esimerkiksi modaalisuutta voidaan monissa kielissä ilmaista niin affikseilla, apuverbeillä kuin erillisillä partikkeleilla. Merkitys- ja funktiolähtöisessä lähestymistavassa näitä kaikkia käsitellään samassa luvussa, vaikka ne kieliopillisen muodon kannalta edustaisivatkin erilaisia rakenteita. Funktiolähtöistä kielioppia kirjoitettaessa pyritään korostamaan erityisesti sellaisia kieliopillisia ilmiöitä, jotka ovat keskeisiä tutkittavassa kielessä tai kiinnostavia kieltenvälisestä näkökulmasta. Jos tutkittavasta kielestä löytyy jokin ilmiö, jota on tutkittu kielitypologian piirissä vain vähän (vaikkapa harvinaisia tiedonlähteiden tyyppejä sisältävä evidentiaalisysteemi), se todennäköisesti kiinnostaa tutkijoita ja sitä kannattaa tuoda esiin.

Useimmat kieliopit ovat rakenteeltaan yhdistelmä muoto- ja funktiolähtöisiä lukuja. Kielioppi aloitetaan usein muutamalla muotolähtöisellä luvulla (esim. äänneoppi, sanaluokat) ja näiden jälkeen tulee painotukseltaan funktiolähtöisempiä lukuja (esim. valenssi, omistus, aspekti). Jos

tutkittavassa kielessä on paljon sanaluokkia, jotka voidaan selkeästi erottaa toisistaan niille ominaisten muotokategorioiden perusteella, kielenkuvaus on usein järkevää rakentaa sanaluokkien ja niille ominaisen morfologian ympärille. Tällaisia kieliä ovat monet indoeurooppalaiset kielet, kuten indoarjalainen rājbanshin kieli. Wilden (2008) rājbanshin kielioppi perustuu siksi pitkälti eri sanaluokkia (substantiivit, pronominit, verbit, apuverbit, adjektiivit, partikkelit, konjunktiot) ja niille ominaisia taivutuskategorioita käsitteleviin lukuihin. Toisaalta taas kielioppi on mahdollista perustaa hyvin pitkälti kieliopillisten merkitysten ja funktioiden kuvaamiseen, jolloin muotokategorioita käsittelevät luvut jäävät paljon vähäisempään rooliin (esimerkkeinä hyvin funktiolähtöisesti kirjoitetuista referenssikieliopeista ks. Frajzyngier 2002 Pohjois-Kamerunissa puhuttavasta hdin kielestä ja Epps 2008 Amazonin alueella puhuttavasta hupin kielestä). Edellä mainittu lähestymistapa toimii erityisen hyvin sellaisen kielen kuvaamiseen, jossa eri sanaluokille ominaiset muotokategoriat ovat vähemmän selkeitä.

Samaa kieliopillista rakennetta voidaan myös käsitellä eri luvuissa sekä muodon että merkityksen näkökulmista. Jos tutkittavassa kielessä on vaikkapa akkusatiivisijaa ilmaiseva pääte, sitä voidaan ensin käsitellä morfologisen rakenteen kannalta substantiivien muotokategorioita käsittelevässä luvussa ja sen merkityksiä ja funktioita objektin merkitsijänä voidaan myöhemmin käsitellä tarkemmin argumenttien merkintää käsittelevässä luvussa.

Muoto- ja funktiolähtöinen esitystapa palvelevat osin erilaisten kohdeyleisöjen tarpeita. Esimerkiksi tarkasteltavan kielen lähisukukieliä tunteva tutkija voi saada kielestä nopeammin kokonaiskuvan muotokategorioittain kuvatun kieliopin avulla. Valtavaa määrää eri kielten kielioppeja aineistonaan käyttävän kielitypologin näkökulmasta taas merkitykset ja funktiot ovat helpoin lähtökohta. Merkitys- ja funktiolähtöinen kieliopin kuvaustapa auttaa kieliopin käyttäjää löytämään haluamansa tiedon, koska samaa kieliopillista ilmiötä usein ilmaistaan monilla erilaisilla rakenteilla. Jos kiinnostuksen kohteena on vaikkapa modaalisuus, kausaalisuus tai omistus, on tiedon hakeminen huomattavasti helpompaa, mikäli koko ilmiötä on käsitelty yhdessä merkityslähtöisessä luvussa ja käyttäjä saa sitä tarkastelemalla kokonaiskuvan aiheesta.

## 2.4. Kieliopin kuvaamisen problematiikkaa

Kieliopin kuvauksessa kaikki liittyy kaikkeen eikä yksittäisiä ilmiöitä ole useinkaan mahdollista kuvata ottamatta samalla kantaa koko kielioppiin. Esimerkiksi sanaluokat voidaan määritellä vain suhteessa toisiinsa. Kun kuvataan substantiiveille ominaisia morfologisia ja syntaktisia piirteitä, otetaan samalla kantaa siihen, mikä erottaa substantiivit verbeistä, ja silloin on tiedettävä jotakin myös siitä, miten verbit kielessä toimivat. Koska kielioppi koostuu lukuisista toisiinsa liittyvien muotojen ja merkitysten alasysteemeistä, kieliopin kuvausta ei ole mahdollista kirjoittaa valmiiksi luku kerrallaan, vaan työ vaatii useiden lukujen työstämistä samanaikaisesti. Jo työn alussa on tärkeää pyrkiä hahmottamaan jonkinlainen kokonaisrakenne kieliopista ja laatia alustava sisällysluettelo. Kieliopin rakenne muuttuu ja täydentyy vähitellen työskentelyn edistyessä, mutta alustava sisällysluettelo on välttämätön tutkimusaiheen kokonaisuuden hahmottamiseksi. Jotkin kieliopin osa-alueet ovat työn kokonaisrakenteen kannalta tärkeämpiä kuin toiset. Tällainen ilmiö on esimerkiksi sanaluokat (Munro 2006, 71). Jos minkä tahansa kielen sanaluokkajärjestelmä analysoidaan uudella tavalla kirjoitusprosessin kuluessa, tämä todennäköisesti aiheuttaa eriasteisia muutoksia kieliopin muihin lukuihin. Toisaalta joitakin kieliopin osa-alueita voi olla mahdollista tarkastella erikseen ilman, että koko työn rakenne radikaalisti muuttuu. Jo kieliopin kuvauksen alkuvaiheessa olisi pyrittävä hahmottamaan, millaisten ilmiöiden kuvaus vaikuttaa eniten kieliopin kokonaisrakenteeseen. Kieliopin kirjoittaja joutuu myös aina tekemään ratkaisuja kuvauksen kattavuuden ja tiiviin, käyttäjäystävällisen esitystavan välillä. Referenssikieliopin täytyy kuvata tutkittava kieli riittävän laaja-alaisesti ja syvällisesti, mutta samalla raja on vedettävä johonkin. Liian laaja ja loputtomia yksityiskohtia sisältävä kielioppi ei enää ole käyttäjäystävällinen eikä helppolukuinen.

Kieliopin luettavuuteen voidaan vaikuttaa monilla kirjoitusteknisillä ratkaisuilla. Vaikka samoja asioita joudutaan monesti käsittelemään useissa kohdissa kielioppia, liiallista toistoa pyritään välttämään ja käytetään sen sijaan tekstin sisäisiä viitteitä, joiden avulla lukijaa ohjataan löytämään kaikki ne luvut, joissa käsiteltävä asia (esim. akkusatiivisija)

esiintyy. Myös lähdeluettelon ja kieliopin loppuun tulevan asiasanahakemiston avulla autetaan lukijaa löytämään haluamansa tieto kieliopista.

Kieliopeissa pyritään käyttämään sellaista terminologiaa, joka on mahdollisimman hyvin tunnettua potentiaalisen kohdeyleisön keskuudessa; termivalinnat voivat tällöin olla hyvinkin erilaisia riippuen siitä, kirjoitetaanko ensisijaisesti tutkijoille vai kieliyhteisölle. Jos termivalinnat eivät ole kohdeyleisölle tuttuja, lukija ei välttämättä löydä haluamaansa tietoa nopeasti katsomalla sisällysluetteloa. Vaikkapa monissa mandariinikiinan kieliopeissa yleinen termi *disposal construction* on tuttu kaikille kiinalaisiin kieliin perehtyneille, mutta muunlaisesta taustasta tuleva kielitypologi ei todennäköisesti tiedä, että kyseinen termi viittaa kiinalaisten kielten kuvaustraditiossa semanttisten ja pragmaattisten seikkojen säätelemään objektin merkintään. Jälkimmäiselle kohderyhmälle termi *Differential Object Marking* olisikin paljon tutumpi. Kuitenkin kieliopissa kuvattujen ilmiöiden huolellinen selittäminen ja käytettyjen termien määrittely on vielä tärkeämpää kuin terminologiset valinnat. Kielioppia aineistonaan käyttävien typologien työtä helpottaa, jos kielioppiin (esim. hakemistoon) merkitään myös, mikäli jotakin keskeistä kategoriaa ei löydy kielestä. Kaikkia mahdollisia kategorioita ei tietenkään voida luetella, mutta jos jokin tutkittavan kielen sukukielille tai saman alueen kielille erityisen tyypillinen rakenne puuttuu tukittavasta kielestä, tästä olisi hyvä erikseen mainita. Samoin jos jostakin kieliopin ilmiöstä ei vielä ole riittävää tietoa, tämä tulee mainita kieliopissa sen sijaan, että ilmiö jätettäisiin kokonaan käsittelemättä. Hyviä ohjeita kieliopin kirjoittamiseen kohdeyleisön näkökulmasta sisältää esimerkiksi Noonan (2006).

## 3. Kielenkuvaushankkeen suunnittelu

Kielenkuvaushankkeen kohdekieli on yleensä tutkijalleen ennalta ainakin osittain tuntematon, ja kenttätyö tapahtuu tutkijan kannalta monelta osin vieraassa ympäristössä. Niinpä huolellinen valmistautuminen

etukäteen on onnistuneen kenttätyömatkan edellytys. Tässä luvussa keskitymme valmistelutyöhön, jota onnistunut kielenkuvaushanke ja kenttätyö edellyttävät. Luvussa 3.1. tuomme esiin kenttämatkan etukäteissuunnitteluun liittyviä käytännön asioita tutkijan kannalta. Luvussa 3.2. tarkastelemme kenttätyötä tutkijoiden ja kielenoppaiden yhteistyön näkökulmasta.

## 3.1. Kenttätyön käytännön valmistelut

Kenttätyön suunnitteluun vaikuttavat usein huomattavasti itse tutkimustilanteen ulkopuoliset tekijät. Eräs tärkeimmistä seikoista on kohteena olevan alueen poliittinen ja turvallisuustilanne. Lisäksi joissakin maissa (esim. Pohjois-Venäjällä tai Länsi-Kiinassa) tietyt alueet voivat olla ulkomaalaisilta suljettuja tai niillä liikkumiseen vaaditaan erikoislupa. Alueelle, jolla ulkomaalaisten liikkumista on rajoitettu, on ehkä mahdollista matkustaa vain muutamaksi viikoksi tai kuukaudeksi kerrallaan, eikä usean kuukauden mittainen kenttätyö näin ole mahdollista. Moniin maihin on haettava viisumi ennen matkaa, ja joissakin maissa saatetaan viisumin lisäksi vaatia erillisiä lupia paikallisviranomaisilta. Esimerkiksi Venäjälle tai Kiinaan suuntautuvalla kenttätyömatkalla tutkijoiden täytyy ennen matkaa anoa viisumit ja kohdealueelle saapumisen jälkeen vielä rekisteröityä paikalliselle viranomaiselle. Hotellit hoitavat rekisteröinnin automaattisesti, mutta mikäli tutkijat majoittuvat esimerkiksi paikallisen perheen luo, on rekisteröitymisestä huolehdittava itse. Poliittisen tilanteen lisäksi kenttätyön suunnitteluun vaikuttavat myös esimerkiksi paikalliset luonnonolot, kulkuyhteydet ja majoitusmahdollisuudet. Esimerkiksi vaativissa luonnonolosuhteissa, kuten korkealla vuoristossa, tai vaikeiden kulkuyhteyksien päässä sijaitseva matkakohde asettaa työskentelylle omat haasteensa. Kulkemista syrjäisimpiin kohteisiin tai majoittumista niissä ei välttämättä ole mahdollista järjestää etukäteen, vaan paljon voi jäädä järjestettäväksi paikan päällä.

Myös tutkijoiden oma elämäntilanne, terveydentila, kielitaito, persoonallisuus ja henkilökohtaiset mieltymykset vaikuttavat kenttätyön suunnitteluun, aina valitusta kohteesta majoitusratkaisuihin ja

työskentelytavoista kenttätyön kestoon. Riittävä kielitaito kohdealueella työskentelyyn ja käytännön asioiden hoitamiseen on onnistuneen kenttätyön edellytys. Monissa maissa puhutaan laajalti jotakin suurta eurooppalaista kieltä, jonka taito on suomalaistutkijoiden keskuudessa yleistä (esim. englanti, ranska, espanja), mutta on myös alueita, joilla työskentelyyn tarvitaan hieman harvemmin osattuja kieliä (esim. Kiinassa mandariinikiinaa, Papua-Uuden-Guinean alueella tok pisiniä). Taatakseen onnistuneen yhteistyön kieliyhteisön kanssa ammattitaitoinen tutkimusryhmä perehtyy etukäteen myös paikalliseen kulttuuriin, kuten tapoihin, uskontoihin, kohteliaisuusnormeihin, alkoholinkäyttötottumuksiin ja sukupuolirooleihin. Paikallisten tapojen kunnioittaminen on kenttätyötä tehdessä ehdottoman tärkeää, ja on hyvä tiedostaa, että esimerkiksi länsimaissa hyväksytyt tavat pukeutua tai käyttää alkoholia eivät ole yhtä hyväksyttyjä kaikissa kulttuureissa. On tärkeää, että kenttätyötä suunnitteleva ryhmä kykenee työskentelemään pitkäjänteisesti kotimaastaan poikkeavissa olosuhteissa ja myös viihtymään kentällä ja nauttimaan työskentelystä.

Kenttätyömatkan valmisteluihin kuuluvat luonnollisesti myös matkabudjetin laatiminen ja rahoituksen järjestäminen, matkojen ja majoitusten järjestäminen, kohdealueen tartuntatautitilanteen selvittäminen ja tarvittavien rokotusten hankkiminen sekä kenttämateriaalin keruussa ja matkalla tarvittavien materiaalien ja tarvikkeiden hankinta ja testaaminen. Välttämättömiä ennakkovalmisteluja ovat myös nauhoitus- ja tutkimustilanteiden sekä tehtävien suunnittelu, joista kerrotaan enemmän luvussa 4. Lisäksi käydään läpi tutkimusmateriaalin nauhoitukseen tarvittava tekniikka sekä audio- ja videotallenteiden formaatti jatkokäytettävyyden mukaan. Äänitiedostoille esimerkiksi suositellaan pakkaamatonta WAV-tiedostomuotoa, mutta on toki huomattava, että tiedostoformaatit ja vastaavasti suositukset niiden käytöstä uusiutuvat jatkuvasti. Etukäteen on mietittävä myös aineiston varmuuskopiointi kentällä useammalle kovalevylle ja tilanteen salliessa verkkosijaintiin. Ennen kenttätyötä on hyvä suunnitella myös litterointi- ja transkriptiotavat sekä taustatietojen kokoaminen tarvittavia metatietoja varten (ks. luku 5.1. ja 5.2.). Lisäksi kenttätyöhön kuuluu oleellisesti kenttäpäiväkirjan kirjoittaminen päivittäisistä toimista, havainnoista, kysymyksistä,

vastauksista ja oivalluksista. Aineistojen koostamiseen, kuvailutietoihin, litterointiin ja arkistointiin liittyviä asioita käydään tarkemmin läpi luvussa 5.

## 3.2. Yhteisö ja kielenoppaat

Ennen matkaa kenttätyötä suunnitteleva tutkijaryhmä hankkii tavallisesti tutkimuksen kannalta välttämättömiä kontakteja sekä tutkittavan kielen puhujiin että paikallisiin instituutioihin, kuten yliopistoihin tai paikallisviranomaisiin. Esimerkiksi paikalliset tutkijat tai kulttuurityöntekijät ovat usein hyviä yhteistyötahoja, jotka voivat tarvittaessa auttaa etsimään työskentelyyn halukkaita kielenoppaita, tutkijoita majoittavia perheitä ja muita tarvittavia kontakteja. Tutkittavan yhteisön jäsenten on mahdollisuuksien mukaan hyvä olla jo etukäteen tietoisia tutkijan saapumisesta ja halukkaita osallistumaan kielenkuvaushankkeeseen. Usein kielenoppaat valikoituvat kuitenkin sattumien ja kentällä luotujen kontaktien kautta. Lisäksi paikallisten ihmisten voi olla aluksi vaikea ymmärtää kielenkuvaushankkeen tarkoitusperiä, ja tutkijoiden tavoitteet avautuvat heille hiljalleen yhdessä työskentelyn myötä.

Työskentelyn onnistumiseen, aineiston laatuun ja tulosten yleistettävyyteen vaikuttaa suuresti se, keitä hankkeeseen osallistuvat kielenpuhujat ovat eli millaiset ihmiset antavat äänensä kielenkuvaushankkeelle. Tutkimukseen osallistuvia kielenpuhujia kutsutaan *kielenoppaiksi*, joskus käytetään myös termejä *informantti*, *konsultantti* ja *kielimestari*. Työskentelyyn kielenoppaiden kanssa vaikuttaa tutkijoiden ja kielenoppaiden välisen vuorovaikutuksen luontevuus: kuinka helposti syntyy keskustelua, miten hyvin kielenopas lähtee mukaan tutkijan ehdottamiin tutkimusasetelmiin tai kuinka mielellään hän kertoo tutkijan toivomista aiheista. Suuri vaikutus on myös kielenoppaan kielitajulla (ks. Kielitaju kielentutkijan työkaluna tk.) eli sillä, miten hän itse jäsentää kieleen ja sen käyttöön liittyviä ilmiöitä ja miten samalla aaltopituudella hän on tutkijan kanssa. Tutkijan kannalta hyvä kielenopas on luonteva ja rohkea omassa kielenkäytössään, ei arvioi omaa kielitaitoaan liian huonoksi (tämä saattaa olla joskus tilanne monikielisten, valtakieltä

päivittäisenä kielenään käyttävien vähemmistökielen puhujien tapauksessa) eikä pelkää korjata tutkijan tekemiä kielivirheitä ja väärinymmärryksiä. Erilaiset ihmiset voivat sopia kielenoppaiksi erilaisiin tarpeisiin. Esimerkiksi joku, joka ei tunne oloaan lainkaan luontevaksi nauhoitustilanteessa, saattaa olla erinomainen apu tutkijalle muun kerätyn materiaalin purkamisessa ja aineistosta esiin nousseiden ilmausten ja vuorovaikutustilanteiden merkitysten selittämisessä.

Eri puhujat käyttävät kieltä eri tavoin, ja toisaalta samat ihmiset puhuvat eri tilanteissa eri tavoin. Yhden ihmisen idiolekti ei kerro välttämättä siitä, miten muut saman yhteisön jäsenet kieltä käyttävät. Lisäksi kaupassa asioitaessa tai koululuokassa ei välttämättä käytetä kieltä samalla tavalla kuin kotona perheenjäsenten keskuudessa. Aineiston edustavuuden kannalta tutkijoiden on siis otettava huomioon kielen käytön variaatio. Tästä syystä kielenoppaiksi pyritään löytämään eri-ikäisiä ja eri sukupuolia edustavia ihmisiä, mahdollisesti myös monenlaisten ammattien tai sosiaalisten asemien edustajia. Vastaavasti ihanteena olisi tallentaa kieltä erilaisissa vuorovaikutustilanteissa. Kielenoppaan valintaan vaikuttaa kuitenkin myös se, kuinka uhanalainen tutkittava kieli on. Jos kielellä on hyvin vähän puhujia jäljellä, voidaan käytännön syistä joutua työskentelemään esimerkiksi vain yhden kielenoppaan kanssa. Kenttätyötilanteet vaihtelevat suuresti, ja yhtä kaikkiin tilanteisiin sopivaa nyrkkisääntöä kielenoppaiden valinnasta ei ole.

Vaikka onnistunut kenttätyö edellyttää suunnitelmallisuutta ja ennalta mietittyjen tutkimusasetelmien ja haastattelutilanteiden luomista, se on mitä suurimmassa määrin luovaa vuorovaikutusta, ihmissuhdetyötä ja aitoja kohtaamisia. Monesti voi olla tarpeen unohtaa suunnitelmat tehokkaasta työskentelystä ja vain elää ja toimia tilanteen mukana. Kenttätyössä tutkijat matkustavat tavallisesti keskelle yhteisön arkielämää ja kiirettä ja saattavat asua paikallisten ihmisten luona, jolloin heistä on yleensä ylimääräistä vaivaa yhteisölle. Tutkijoiden on siksi tärkeää kunnioittaa paikallisten omaa aikaa ja antaa myös itsestään osallistumalla yhteisön arkeen ja kompensoimalla ylimääräisiä kuluja ja vaivaa. Kenttätutkijan toimenkuvaan voi näin kuulua esimerkiksi halonhakkuuta, lastenhoitoa ja illanviettoja, joissa kaukaa tulleilta vierailta voidaan odottaa esimerkiksi lauluesitystä. Kenttätyö vaatii joustavuutta, kärsivällisyyttä

ja kykyä ymmärtää, että paikallisilla ihmisillä on paljon muitakin velvollisuuksia ja intressejä kielenkuvaushankkeeseen osallistumisen lisäksi.

Kenttätyö sisältää myös suuren tutkimuseettisen vastuun. Tutkijoiden on aina tilanteita ja vuorovaikutusta tallentaessaan huolehdittava, että osallistujat tietävät, mitä nauhoitetaan, miksi ja mihin nauhoitus päätyy. Nauhoittamiseen ja materiaalin jatkokäyttöön on oltava aina osallistujien antama suostumus tai tutkimuslupa, joko kirjallisessa tai suullisesti tallennetussa muodossa. Valmisteluun kuuluu tutkittavien taustatietojen keräämisen ja tutkimusprosessin suunnittelu eettiset kysymykset huomioon ottaen. Tutkijan on etukäteen mietittävä mm. tekijänoikeus- ja yksityisyydensuoja-asioita, suunniteltava lupien kysyminen tutkimukseen ja laadittava mahdolliset suostumuskaavakkeet tutkimusta varten. Osallistujien tulee olla tietoisia siitä, mihin heiltä kerättyjä materiaaleja käytetään, miten niitä käsitellään ja minne aineistot päätyvät. Tutkijan täytyy voida perustella, miksi hänen keräämänsä osallistujia koskevat tiedot ovat tarpeellisia. Pitää muistaa, että esimerkiksi terveydentilaa koskevia tietoja voi tulla esille spontaaneissa nauhoitetuissa keskusteluissa, ja tällaistenkin tietojen käsittely tulee ottaa huomioon osallistujien informoinnissa. On myös tärkeää, että osallistujilla on tutkimuksen aikana mahdollisuus muuttaa mieltään ja vetää pois tuottamansa materiaali.

Tutkimusaineistoon liittyvä eettinen ennakkoarviointi tehdään suunnitteluvaiheessa aineiston koko elinkaaren näkökulmasta. Suostumusten kirjaamisessa ja osallistujien informoinnissa aineiston käyttötapa ja käsittely on tärkeää kuvata niin, että aineisto voidaan myöhemmin luovuttaa tutkimusyhteisön käyttöön myös muihin, eri tieteenalojenkin tutkimustarkoituksiin, mahdollisesti myös opetus-, taide- ja kulttuurityön tarkoituksiin sekä kielenpuhujille itselleen. Yliopistoilla ja tutkimuslaitoksilla, osin myös kieliarkistoilla ja aineistoinfrastruktuureilla, on tutkimusdatan keräämistä, tallentamista, käyttöä ja hallintaa koskevia periaatteita ja linjauksia (esim. tutkimusdatapolitiikka), joihin kannattaa tutustua ennen kenttätyötä. Hyvä peruslähde on Yhteiskuntatieteellisen tietoarkiston Aineistonhallinnan käsikirja.

Lisäksi ennen matkaa on suunniteltava, millä tavoin hankkeeseen osallistuvien kielenoppaiden työ ja aika korvataan. Tutkimusmateriaalin

palauttaminen kieliyhteisölle ja erilaisten kieliaineistojen (esim. oppimateriaali, digitaaliset aineistot) tuottaminen kieliyhteisön hyväksi on tässä isossa roolissa, mutta joissakin tapauksissa myös rahallinen korvaus kielenoppaille tulee kysymykseen. Mikäli kielenoppaille on tarpeen maksaa työskentelystä, sopiva summa on syytä selvittää etukäteen aluetta hyvin tuntevilta tahoilta tai vaikkapa paikallisista palkkataulukoista. Liian pieni korvaus ei ole tietenkään suotavaa, mutta myös liian suurten summien maksamisesta voi aiheutua ongelmia. Ylisuurten palkkioiden maksaminen voi esimerkiksi aiheuttaa eripuraa yhteisön sisällä tai saada aikaan sen, että tutkimukseen osallistuminen ei enää ole kaikille yhteisön jäsenille vapaaehtoista, jolloin kielenoppaat saattavat painostaa muita osallistumaan hankkeeseen saadakseen itse työtä ja rahaa. Liika avokätisyys rahapalkkioissa voi vaikeuttaa myös muiden tutkijoiden toimimista alueella, kun suuriin palkkoihin tottuneet kielenoppaat eivät hyväksy muualta tulevien tutkijoiden maksamia summia (tämä on muodostunut paikoin ongelmaksi esim. Venäjällä). Palkkioiden maksamisessa on huomioitava myös alueelliset säädökset työn maksullisuuden vaikutuksesta tekijänoikeuteen.

# 4. Kenttätyöaineiston keruumenetelmiä

Kielitieteellinen kenttätyö ei välttämättä aina liity kielenkuvaushankkeeseen. Kenttätyötä voidaan tehdä hyvin erilaisia tutkimusasetelmia varten missä tahansa kielenkäyttöympäristössä ja -tilanteessa (ks. myös Lingvistinen etnografia toimintayhteisöissä tk.). Kenttätyötä tarvitaan, kun halutaan tutkia sellaisia ilmiöitä, kielenkäyttötilanteita tai kielimuotoja, joista ei ole saatavilla tutkimustarkoitukseen sopivia valmiita aineistoja, kuten sanomalehti- tai keskustelukorpuksia. Kenttätyö voi liittyä esimerkiksi keskustelujen tutkimiseen asiointitilanteissa tai tietyissä toimintaympäristöissä, paikannimistön keräämiseen jonkin alueen asukkaita haastatellen tai erityyppisten kielen ilmiöiden havainnointiin jostakin vähemmän tutkitusta kielestä. Käytettävät keruumenetelmät riippuvat

aina tutkimuksen tavoitteesta. Koska referenssikieliopin kirjoittamiseen tähtäävä tutkimus edellyttää kielen kaikkien osa-alueiden tarkastelua, siihen liittyvässä kenttätyössä on sovellettava useita keruumenetelmiä. Kenttätyö edellyttää ennalta mietittyjen tutkimusasetelmien ja haastattelutilanteiden luomista, jotta tutkijat saavat kerätyksi tarkoituksiinsa soveltuvan ja käyttökelpoisen materiaalin. Tällaisia menetelmiä ja näkökulmia erilaisiin aineistonkeruutapoihin esittelemme seuraavissa luvuissa.

## 4.1. Luonnollisen puheen tallentaminen

Kenttätutkijan tavoitteena on kielen tallentaminen mahdollisimman luonnollisessa muodossa ja käyttöympäristössä. Luonnollisen vuorovaikutuksen tallentaminen liittyy tyypillisesti pitkäkestoiseen etnografiseen kenttätyöhön (ks. myös Lingvistinen etnografia toimintayhteisöissä tk.), jonka yhteydessä puhutaan usein osallistuvasta vuorovaikutuksen havainnoinnista. Luonnollista vuorovaikutusaineistoa ja kerrontaa voidaan tallentaa myös haastattelutilanteissa. Usein työskentely kielenoppaiden kanssa alkaakin taustatietohaastattelulla, jota seuraavat jatkokeskustelut ja -kertomukset ovat usein arvokkaita nauhoitusaineistoja, joissa kielenoppaat kertovat vapaasti heille läheisistä aiheista. Erityisesti iäkkäämmät kielenoppaat saattavat mielellään kertoa asuinalueensa historiasta, omasta elämästään ja perhehistoriastaan (ks. myös Haastattelu aineistonkeruun metodina sosiolingvistiikassa tk.). Toiset saattavat puhua innokkaasti paikallisista tavoista ja ihmisistä tai ajankohtaisista tapahtumista ja kertoa niihin liittyviä anekdootteja. Paikkoihin liittyviä kertomuksia taas saattaa poikia paikannimistöä keräävien tutkijoiden haastatteluista, joilla pyritään yleensä muodostamaan kokonaiskuva alueen lähiympäristöstä, talojen, metsien, peltojen tai muiden paikkojen sijainneista ja nimityksistä. Asutus- ja paikallishistoriaa käsittelevät keskustelut ja kertomukset toimivat usein monipuolisena aineistona, jota voi kieliopin kuvaamisen lisäksi hyödyntää esimerkiksi kielisosiologisen historian, kielikontaktien tai paikannimistön tutkimuksessa. Perinteisemmin kenttätyössä käytettyjen (kieli)elämäkerta- ja

paikallishistoriahaastatteluiden ja tarinoiden kertomisen lisäksi mitkä tahansa tutkijaa kiinnostavat aiheet, joista kielenoppaat innostuvat, sopivat vapaan puheen nauhoittamiseen.

Vaikka kenttätyön tavoitteista, nauhoitustilanteista ja aineiston käytöstä on tärkeää sopia kielenoppaiden kanssa, tutkijoiden ei välttämättä kannata paljastaa aivan suoraan tutkimiaan kielen ilmiön yksityiskohtia, jotta osallistujat eivät ala kiinnittää huomiota kyseiseen rakenteeseen puheessaan. Vapaata puhetta tallentavien tutkijoiden tehtävänä onkin luoda mahdollisimman luonteva keskustelutilanne, kuunnella ja antaa tilaa puhujille. Seuraavan kysymyksen tai tilanteen ohjaamisen kanssa ei kannata kiirehtiä, vaan on parempi odottaa, jatkaisiko puhuja vielä hetken tuumittuaan. Tämän on yksi jos toinenkin kenttälingvisti oppinut kantapään kautta kuunnellessaan nauhoituksiaan jälkikäteen. Useimmiten tutkijoiden ja tallennusvälineiden läsnäolo kuitenkin vaikuttaa osallistujien kielenkäyttöön, ainakin aluksi (ns. tarkkailijan paradoksista ks. Labov 1972, 209). Puhujat saattavat huomaamattaan muokata tai siistiä puhettaan ja valita ilmaisutapoja, jotka kokevat hienoksi ja hyväksi tavaksi käyttää kieltä. He voivat nauhoitustilanteessa esimerkiksi suosia mahdollisesti olemassa olevaa standardisoitua yleiskieltä tai arvostamaansa varieteettia tavallisen puhetapansa sijaan. Joissakin kielissä yhteisön ulkopuolisen ihmisen läsnäolo puhetilanteessa vaikuttaa jopa kieliopillisten muotojen valintaan, kun puhujien ei ole esimerkiksi yhteisön paikallishistoriasta kertoessaan mahdollista käyttää sellaisia kieliopillisia muotoja, joilla viitataan kaikkien läsnäolijoiden jakamaan intersubjektiiviseen tietoon (Bergqvist 2016). Vaikka tutkijan läsnäolon mahdollinen vaikutus on hyvä tiedostaa, monesti myös tutkijan ohjaamien kielenkäyttötilanteiden (stimulukset ja lavastetut kielenkäyttötilanteet, ks. luku 4.2.) voidaan katsoa edustavan suhteellisen luonnollisesti tuotettua kieltä.

Kielen dokumentointia varten tehdyt nauhoitukset puretaan tavallisesti jollakin tasolla jo kenttätyökohteessa, sillä niiden sisältöjen ja rakenteiden täsmällinen ymmärtäminen vaatii usein äidinkielisen puhujan apua. Kaikki vapaata puhetta sisältävät nauhoitukset onkin hyvä käydä läpi yhdessä kielenoppaiden kanssa, tapahtui se sitten kenttätyökohteessa heti, myöhemmin nauhoitustilanteen jälkeen tai paljon

myöhemmin internetyhteyden välityksellä, esimerkiksi videopuheluiden avulla. Nauhoituksissa esiintyvistä rakenteista ja ilmiasuista myös herää usein lisäkysymyksiä, joiden tutkimusta on hyvä päästä jatkamaan kenttätyön aikana muiden metodien avulla. Kun tutkimuksen tarkoituksena on kattaa kaikki tai useimmat kieliopin osa-alueet, vapaata puhetta sisältävät nauhoitukset eivät yksin riitä aineistoksi. Tallennetuissa kertomuksissa ei välttämättä tule esiin vaikkapa käskymuotoja tai kysymyslauseita. Toisaalta tietyt kieliopilliset muodot, ilmaisut tai rakenteet saattavat olla arkipuheessa harvinaisia, tai ne voivat ilmetä vain rajatuissa tilanteissa, ja niiden merkitykset saattavat riippua puhetilanteen osallistujista. Joitakin tutkimuskysymyksiä varten tarvitaan muita menetelmiä, erityisesti lavastettuja kielenkäyttötilanteita, joissa tutkijalla on mahdollisuus kontrolloida, muuttaa ja tarkistaa osallistujien puheen välittömään kontekstiin liittyviä tekijöitä. Näitä menetelmiä käsitellään seuraavaksi.

## 4.2. Stimulukset ja lavastetut kielenkäyttötilanteet

Kun kenttätutkija ohjaa tavalla tai toisella kielenkäyttötilannetta, puhutaan elisitaatiotekniikoista. Tällaisia voivat olla virikkeet eli *stimulukset*, joiden tarkoitus on antaa osallistujille puheenaiheita ja ohjata heitä käyttämään kieltä valikoiduissa tilanteissa tai puhumaan tutkijan toivomista aihepiireistä. Virikkeitä voidaan käyttää yhtä lailla yksilö- tai ryhmähaastatteluissa, tutkimusasetelman tarpeen mukaan. Niiden tarkoitus on myös saada puhujat käyttämään kieltä luonnollisesti ilman toisen kielen vaikutusta, kuten tapahtuisi käännettäessä valikoituja lauseita ja sisältöjä suoraan yhteisestä metakielestä tutkitulle kielelle. Kenttätyössä käytettäviä virikkeitä voivat olla piirrokset, valokuvat, kuvasarjat, videot ja esineasetelmat. Kielten kuvaushankkeissa paljon käytettyjä apuvälineitä ovat esimerkiksi päärynätarina (engl. *Pear story*[4], ks. Chafe 1980) ja sammakkotarinat (engl. *Frog stories*[5]). Sammakkotarinat ovat kuvasarjoina, päärynätarinat videopätkinä esitettyjä lyhyitä juonellisia kertomuksia, joiden pohjalta kielenoppaiden tehtävänä on kertoa tarina kielellisessä muodossa. Virikkeenä voi toimia myös jonkin tutun asian valmistamista

tai muuta askaretta kuvaava lyhyt video, jonka tapahtumat tutkija pyytää kielenpuhujaa kertomaan järjestyksessä. Tarkkaan strukturoidussa elisitaatiotilanteessa tehtävänä voi myös olla lyhyen reitin neuvominen paikasta A paikkaan B kartan tai reittikuvan pohjalta.

Paljon käytettyjä virikkeitä ovat myös erilaiset esineet, esimerkiksi nuket, ja niistä muodostettavat muuttuvat asetelmat, joista kielenoppaita pyydetään selittämään esineiden suhteita ja niiden kuvaamista tapahtumissa. Kuvalliset tehtävät ja apuvälineinä käytettävät esineet toimivat hyvin virikkeinä konkreettisten ilmausten tutkimisessa. Tällaisia voivat olla vaikkapa värinnimitysten tunnistaminen kuvista, tekemistä kuvaavien tapahtumien (kuten rikkoa, hypätä, kiivetä) esittäminen kuvasarjoina tai videoina tai spatiaalisten suhteiden ilmausten (kuten edessä, taakse, alta) elisitoiminen leluasetelmien avulla. Hyviä esimerkkejä esineasetelmien käytöstä tilan ja sijainnin sekä tapahtumien ilmausten tutkimisessa ovat Hellwig (2006) afroaasialaisen goemain kielen lokatiiviverbien tutkimuksesta, Brown (1993) mayakieli tzeltalin absoluuttisen spatiaalisen systeemin tutkimuksesta muovipaimenen ja muovipuun avulla sekä Benjamin Brueningin kehittämä the Scope Fieldwork Project -materiaali[6]. Erilaisia elisitaatiotekniikoita ja niiden soveltamista ovat kuvanneet yksityiskohtaisesti muun muassa Chelliah ja de Reuse (2011). Bouquiaux ja Thomas (1992) puolestaan sisältää runsaasti eri tarkoituksiin soveltuvia kyselylomakkeiden mallipohjia sekä piirroksia, joita voidaan käyttää erikoisalojen sanaston (esim. kasvillisuuteen ja eläimistöön liittyvän terminologian) keräämisen apuna. Kielten dokumentointiin ja kenttätutkimukseen soveltuvia materiaaleja on koottu tutkijoiden ladattaviksi myös eräiden typologiaan ja kielten dokumentointiin keskittyneiden instituuttien verkkosivuille[7].

Stimulusten on hyvä olla kulttuurikontekstiin sopivia: jotkin aiheet voivat olla joissakin kulttuureissa tabuja, esim. kuvat tietynlaisesta vaatetuksesta tai sosiaalisesta käytöksestä voidaan kokea loukkaavina. Lisäksi materiaalin olisi kuvattava puhujille tuttua fyysistä ympäristöä ja tuttuja esineitä (ks. esim. Himmelmann 1998, 29). Päärynätarina ei sovi parhaalla mahdollisella tavalla pohjoissiperialaiseen kontekstiin, eivätkä muovileluja kuvaavat valmiit kuva-asetelmat tai sarjakuvat välttämättä avaudu kaikissa kulttuureissa. Paikannimistöä keräävän tutkijan työtä

voi hyödyttää suuresti, jos hänellä on mahdollisuus katsoa karttaa yhdessä kielenoppaan kanssa, mutta kartanlukutottumus tai kyky tulkita karttojen ja kuvien mittakaavaa ei ole kaikkialla itsestäänselvyys. Toisaalta kielenoppaat saattavat hyvinkin luovasti osata korvata stimulusmateriaalissa esiintyviä omalle kulttuurilleen vieraita asioita tutummilla, jolloin kuvakortissa näkyvä lehmä korvautuu spontaanisti porolla, ananas omenalla ja palmu koivulla, tai toisin päin.

Vähemmän ohjattu tutkimustilanne on sellainen, jossa tutkija voi näyttää tai esittää osallistujille jonkin kuvan, videon tai tekstin ja pyytää keskustelemaan sen pohjalta vapaammin. Tällaisia tutkimusasetelmia ovat lavastetut tai järjestetyt kielenkäyttötilanteet, joissa osallistujat asetetaan ennalta suunnitellun tehtävän äärelle. Tällaisilla tehtävillä voidaan pyrkiä luomaan kielenkäyttötilanteita, joissa esimerkiksi kielen ja sen rakenteiden pragmaattiset funktiot tulevat esille ja joissa niiden käyttöön liittyviä kontekstimuuttujia voidaan tarkastella. Tarkasti suunniteltu mutta hyvin vapaata puhetta tuottava tutkimusasetelma on esimerkiksi sellainen, jossa kielenoppaiden tehtävänä on (yhdessä tutkijoiden kanssa) kävellä heidän yhteisölleen tai ympäristölleen merkittävä maastoreitti ja selostaa matkan varrella liikkumistaan ja vastaan tulevia kohteita. Australian alkuperäiskielten tutkijat ovat esimerkiksi keränneet tällaisia maastokävelyiden reittiselostuksia, joita he ovat pystyneet yhdistämään tarkkoihin topografisiin ja sijaintitietoihin kielenoppaiden kantamien GPS-koordinaatteja ja videokuvaa tallentavien kameroiden avulla. Aineistoa on kerätty spatiaalisten suhteiden, suunnan ja liikkeen ilmausten sekä paikannimistön tutkimiseksi. (Cialone 2017.)

Modaaliset ja evidentiaaliset rakenteet sekä erityisesti niiden vuorovaikutukselliset funktiot ovat esimerkki kielen osa-alueista, joita ei yleensä ole helppo tavoittaa nauhoitetuista kertomuksista tai kaikista keskusteluista. Niiden nauhoittamiseen ja elisitointiin on kuitenkin kehitetty hyvin toimineita tekniikoita. Esimerkiksi Silva ja AnderBois (2016) ovat tutkineet Etelä-Amerikassa puhuttavan desanon kielen evidentiaalisia ja episteemisiä (päättelyä ja todennäköisyyttä merkitseviä) ilmauksia käyttämällä apuna Mastermind-lautapeliä. Logiikkapelin tarkoituksena on päätellä eriväristen nappuloiden järjestyksestä muodostettu oikea värikoodi. Tutkijat pyysivät kielenoppaitaan pelaamaan peliä yhteistyössä

ja neuvottelemaan mahdollisuuksista ja vaihtoehdoista ja saivat siten nauhoitetuksi runsaasti päättelyä, epävarmuutta, todennäköisyyttä ja mahdollisuutta ilmaisevia rakenteita luontevissa käyttötilanteissa. San Roque ja muut (2012) puolestaan ovat kehittäneet Family Problems Picture Task -nimisen 16 piirroskuvan kokoelman, jossa perheenjäseniksi helposti tulkittavissa olevat ihmishahmot esiintyvät monenlaisissa tilanteissa, joissa hahmojen yllä olevien kuvitettujen puhekuplien avulla kuvataan ristiriitoja, tunteita ja dramatiikkaa. Tehtävään osallistuvien kielenoppaiden on tarkoitus samaistua hahmoihin ja tuntea ja ilmaista empatiaa näitä kohtaan, mistä syystä kuvasarja on toiminut hyvin kognitiivisia toimintoja kuvaavien ilmausten ja kategorioiden elisitoinnissa.

## 4.3. Käännöstehtävät ja kieliopillisuusarviot

Kielestä toiseen kääntäminen, niin sanottu käännöselisitaatio, ei koskaan ole ensisijainen tai suositeltava keino sen selvittämiseksi, miten kuvattavassa kielessä jokin asia ilmaistaan. Yksittäisten sanojen tai hyvin yksinkertaisten ilmausten kääntäminen on todennäköisesti ongelmatonta, mutta hiukankin monimutkaisemman ilmauksen tai lauseen kääntämisessä on aina olemassa se riski, että kielenoppaat tuottavat tutkijan tarjoilemasta lauseesta sanatarkan käännöksen, joka ei kuitenkaan ole idiomaattinen tapa ilmaista asiaa heidän omassa kielessään. Kontekstualisoitu elisitaatio, jossa tutkija antaa kielenoppaalle jonkin tilanteen tai tapahtuman, johon kielenopas täydentää haetun ilmauksen, toimii usein paremmin kuin pelkkä kääntäminen.

Joissakin tilanteissa, esimerkiksi tutkittavan kielen taivutusparadigmoja kerättäessä, suoraan toisesta kielestä kääntäminen on kuitenkin hyvä menetelmä. Käännöslauseiden elisitoinnilla voidaan täydentää tai tarkistaa kielen kuvausta varten joitakin keskustelussa ja kertomuksissa harvemmin esiintyviä muotoja. Tällainen voisi olla vaikkapa subjekti- ja objektikongruenssia merkitsevä taivutuspääte subjektin monikon toiselle persoonalle ja objektin kaksikolle ('Te [sinä, Anna ja Liisa] olette lukeneet nuo kirjat [2 kpl]'). Myös esimerkiksi kielen foneemien erottelussa suora kääntäminen voi olla nopea keino tarkistaa minimipareja eli

sanapareja, jotka eroavat toisistaan vain yhden äänteen kohdalla. Käännösten ohella toinen hyvin suoraviivainen ja siksi myös varovaisesti käytettävä tutkimusmetodi on tutkijan tarjoamien muotojen ja ilmausten kieliopillisuuden arvioituttaminen kielenoppaalla. Hyväksyttävyystestit ("Voiko näin sanoa, onko tämä oikein?") voivat olla hyvin oleellisia kieliopin kokonaiskuvauksen kannalta, jolloin myös negatiivinen evidenssi on tärkeää. Se, onko joidenkin kieliopillisten muotojen yhdistäminen kielessä mahdotonta vai eikö yhdistelmä vain ole esiintynyt tutkijan aineistossa, selviää yleensä parhaiten hyväksyttävyystesteillä. Esimerkiksi suomen kielen potentiaali ei useimpien puhujien kielitajun mukaan toimi imperfektissä (*Pohtinen tätä tänään* mutta **Pohtinin tätä eilen*).

Kielentutkijan tarkoitusta hyvin ymmärtävän ja kielioppiin analyyttisesti suhtautuvan kielenoppaan kanssa voi olla helppoa käydä joitakin asioita läpi käännöslauseiden ja hyväksyttävyysarvioiden avulla, mutta on muistettava, että kuten tutkijankin oma kielitaju myös informantin kielitaju saattaa pettää. Esimerkiksi kohteliaisuus- tai arvotussyistä kielenopas saattaa vakuuttaa tutkijalle (ja itsekin uskoa) ilmaisevansa jonkin asian toisella tavalla kuin tositilanteessa sanoisi. Työskennellessä on tärkeää huomioida myös kielenoppaan jaksaminen ja väsymyksen vaikutus. Kielenopas voi väsyneenä esimerkiksi hyväksyä epäkieliopillisia muotoja joko epähuomiossa tai siksi, ettei enää jaksa korjata niitä. Liian pitkiä työskentelysessioita ja useita samanlaisia peräkkäisiä tehtäviä on siis vältettävä, jotta keskittyminen pysyy yllä. On myös muistettava, että joissakin tapauksissa elisitointitehtävät voivat olla ainoa mahdollinen aineistonkeruumenetelmä. Luonnollisen kielen nauhoittaminen ei ole aina mahdollista silloin, kun kyseessä on hyvin vakavasti uhanalainen kieli. Tällöin kieltä vielä osaavien ihmisten kielitaito voi olla puutteellinen, kun he eivät enää käytä tutkittavaa kieltä aktiivisesti arkielämässään.

**Sanastoa**

Annotaatio: puheen tekstimuotoon saattamisessa käytettävä merkintätaso, kuten litteraatti, käännös, sanasto- ja morfeemianalyysi.

Elisitaatio (kirvoitus, engl. *elicit* 'saada esiin'): tutkijan käyttämä kenttätyöaineiston keräystekniikka.

Glossaus: kolmelle riville merkittävä morfeemianalyysi, jossa erotetaan kielenaineksesta sana- ja morfeemirajat, nimetään morfeemit ja merkitään käännös.

Kielenopas (informantti, konsultantti, kielimestari): kenttätyössä haastateltava tutkittavan kielen puhuja.

Litterointi: nauhoitteen kirjoittaminen tekstimuotoon.

Metadata eli metatiedot: aineiston yhteyteen tallennettavat taustatiedot kenttätyötilanteesta.

Stimulus (virike): apukeino, jolla pyritään herättämään ja ohjaamaan puhetta. Esim. video, peli, kuva, esine.

Transkriptio: deskriptiivinen puheen kirjoitustapa, jolla puhetta voidaan kuvata foneettisesti tai fonologisesti.

# 5. Kenttätyökorpuksen koostaminen: materiaalista aineistoksi

Kentällä kerätyn audio-, video- ja muun materiaalin työstäminen tutkimusaineistoksi ja kielen dokumentointihankkeen tulokseksi vaatii aina huolellista jatkokäsittelyä. Sekä kenttälingvistin omaa tutkimusta että muuta myöhempää käyttöä varten aineiston tulee olla hyvin järjestetty ja merkitty käyttötarkoitukseen sopivilla tavoilla, ja se tulee valmistella pysyvää säilytystä ja yleensä myös julkaisua silmällä pitäen. Aineiston työstämiseen sisältyvät yleensä seuraavat työvaiheet:

materiaalin käsittely jatkokäytön ja siirrettävyyden mahdollistavilla sovelluksilla ja tallennus ajallisesti kestäviin, siirrettäviin ja arkistointikelpoisiin formaatteihin, aineistoa kuvailevan taustatiedon eli metadatan tuottaminen, julkisuus- ja käyttöoikeusluokituksista päättäminen, aineiston osien litterointi tekstimuotoon, käännökset, eritasoiset sisältömerkinnät sekä arkistointi jatkokäytön ja pysyvän säilytyksen takaamiseksi.

## 5.1. Aikakohdennetut annotaatiot

Digitaalisesti tallennetun aineiston järjestelmällistä merkintää ja muistiinpanoin varustamista kutsutaan yleensä *annotoimiseksi*[8]. Kieliaineistoihin sovellettavia merkintä- eli annotaatiotasoja ovat tyypillisesti ainakin puheaineiston tai sen osien litterointi eli puheen sisällön kirjoittaminen tekstimuotoon sekä käännös aineiston käyttäjien kannalta tähdelliselle laajemmin ymmärretylle kielelle. Litteroinnin ja käännöksen lisäksi merkitään usein aineiston laadun ja käytön kannalta tarvittavia annotaatiotasoja, kuten lause-, sana- tai morfeemitason analyysit, vuorot, eleet ja puheen sisältöä koskevat merkinnät. Kieliopin kuvaamista varten tehdään tavallisesti litterointi sekä lause-, sana- ja morfeemitason annotaatiot. Aineiston käyttötarkoitus määrää sen, mitä osia materiaalista annotoidaan ja millä tavoilla.

Litteroinnissa valitaan sellainen kirjoitusjärjestelmä eli transkriptio, joka yhtäältä palvelee parhaiten aineiston ensisijaista käyttötarkoitusta, toisaalta on ymmärrettävä ja helposti tulkittava mahdollisimman monipuoliselle aineiston käyttäjäkunnalle. Jos tutkimuskohteena olevalla kielellä on oma puhujayhteisön käyttämä ortografia, voidaan sitä joko sellaisenaan tai muokattuna, esimerkiksi kyrilliseltä aakkostolta latinalaisille kirjaimille translitteroituna, käyttää litteroinnissa. Kielelle, jolla ei ole omaa kirjoitusjärjestelmää, käytetään aina kielitieteelliseen tutkimukseen tarkoitettuja transkriptioita. Ennen äänitallenteiden aikakautta foneettinen transkriptio eli äänteiden hyvin täsmällinen merkintä tarkekirjoituksella oli ensiarvoisen tärkeää kenttätyöaineistoissa. Nykyisin litteraatit voivat olla karkeampia, mutta tyypillistä on pyrkiä kuvaamaan transkriptiolla kielen foneemisto. Transkriptioiden tarkkuus riippuu

kuitenkin lopulta aineiston käyttötarkoituksesta. Kirjaimistoista nykyisin laajimmin käytössä on kansainvälinen foneettinen kirjaimisto IPA (*International Phonetic Alphabet*), jonka avulla voidaan merkitä kielten äänteitä ja foneemeja. Eri kielten ja kieliryhmien kirjoittamiseen käytetään myös muita kirjoitusjärjestelmiä, esimerkiksi suomalais-ugrilaista tarkekirjoitusta (SUT).

Kieliopin kuvaukseen tarkoitettu aineisto usein *glossataan* eli se esitetään morfologisesti analysoituna. Glossaus on kielioppien lause-esimerkeissä ja kielennäytteissä käytetty esitystapa, joka kertoo kieliopillisen perusrakenteen. Glossauksessa litteroituun kielenainekseen merkitään morfeemirajat, ja alemmalla sanarajojen mukaan kohdistetulla rivillä kerrotaan tarkka käännös kullekin sanavartalolle sekä sanamuotoon merkittyjen kieliopillisten kategorioiden lyhenteet kapiteelikirjaimilla.

1)

| | | | | |
|---|---|---|---|---|
| *Tämä* | *on* | *esimerkki* | *glossa-tu-sta* | *lausee-sta.* |
| this | be.3SG | example | gloss-PTCP.PST.PASS-ELA | sentence-ELA[9] |

’This is an example of a glossed sentence.’

Tarkempia ohjeita glossaukseen antavat Leipzigin glossaussäännöt (*Leipzig Glossing Rules*[10]). Jotta glossiin merkitty rakenneanalyysi olisi helposti luettavissa myös analysoitua kieltä osaamattomalle tutkijalle, käytetään mahdollisuuksien mukaan vakiintuneita lyhenteitä, joista on tiivis lista Leipzigin sääntöjen yhteydessä; esimerkiksi lokatiivi lyhennetään LOC ja mennyt aika PST (*past tense*). Kielikohtaisille kategorianimityksille on kuitenkin syytä käyttää niitä kuvaavia lyhenteitä (esim. INE = inessiivi, ELA = elatiivi jne). Tärkeintä on, että käytetyt lyhenteet selitetään.

Yksi tärkeimmistä digitaalisten apuvälineiden tuomista hyödyistä aineistonhallinnalle on litteraatin ja muiden annotaatiotasojen aikakohdistaminen eli synkronointi nauhoitteen vastaavien osioiden kanssa. Aikakohdennetussa aineistossa audio- tai videonauhoitteen tietyn jakson valitsemalla saa samalla esille siihen kohdennetut annotaatiotasot.

Tätä tarkoitusta varten on olemassa useita avoimeen lähdekoodiin ja xml-formaattiin perustuvia työkaluja. Kenttätyöaineiston annotointiin ja kielten dokumentointiprojekteihin hyvin soveltuva ohjelma on Elan[11], joka tukee sekä video- että audioaineistoja ja on tästä syystä käytössä myös viittomakielisten aineistojen käsittelyssä. Annotaatio-ohjelman valinta riippuu kuitenkin aina siitä, millaiset merkintätasot ovat tärkeitä. Esimerkiksi puheanalyysiin ja foneettiseen tutkimukseen soveltuu hyvin Praat-ohjelma[12], joka mahdollistaa akustisen analyysin ja puheen mallintamisen esimerkiksi spektrogrammeilla. Praat ei kuitenkaan tue videotiedostoja, mutta toisaalta se toimii hyvin yhteen Elanin kanssa. Annotointiohjelman valinnassa on otettava huomioon myös aineiston siirrettävyys arkistointiformaatin kannalta. Vanhentuneiden tai hankalasti siirrettävien tiedostomuotojen konvertointiin käytettävän formaattivelan syntymistä on hyvä yrittää välttää suosimalla mahdollisimman laajasti käytettyjä tiedostoformaatteja ja ohjelmistoja.

Koska audio- ja videomateriaalin tallentaminen on usein hyvin helppoa ja annotointi puolestaan erittäin aikaa vievää, koko kerätyn kenttäaineiston tarkka merkitseminen vaatii usein laajahkon dokumentaatioprojektin ja useamman ihmisen työpanoksen. Kieliteknologian tarjoamat mahdollisuudet kehittyvät kuitenkin jatkuvasti, ja automatisoitua annotointia kehitetään myös vähän tutkittujen kielten ja niiden dokumentaation tarpeisiin (ks. esim. Gerstenberger, Partanen, Rießler & Wilbur 2016).

## 5.2. Aineiston arkistointi ja jakaminen

Jotta vaivalla kerätty ja annotoitu kieliaineisto ei unohdu tutkijan omalle koneelle odottamaan kovalevyn tuhoutumista, katoa jäljettömiin tutkijan vaihtaessa alaa tai häviä yliopiston verkkopalvelujen uudistuessa, se arkistoidaan pysyväissäilytystä ajatellen. Tutkimuslaitokset ja instituutit eri puolilla maailmaa tarjoavat palveluja kieliaineistojen digitaaliseen arkistointiin sekä arkistoitujen aineistojen tutkimuskäyttöön. Suomessa kieliaineistojen pysyväissäilytystä ja tutkimuskäyttölupia hallinnoi Kielipankki, josta vastaa suomalaisten yliopistojen ja tutkimuslaitosten

muodostama konsortio. Kielipankin materiaalit ovat osa kansainvälistä aineistoinfrastruktuuria[13], jonka tarkoitus on mahdollistaa materiaalien jakaminen yhteismitallisesti eri arkistointilaitosten ja niiden palveluita hyödyntävien tutkijoiden välillä.

Aineiston arkistointi, jakaminen tutkija- ja kieliyhteisölle ja mahdollinen muu jatkokäyttö on hyvä pitää mielessä jo nauhoituksia ja annotaatioita tehtäessä. Tallenteiden tulee syntyhetkestään saakka olla sellaisessa tallennusmuodossa, joka mahdollistaa niiden siirrettävyyden ja arkistoinnin, joten arkiston mahdolliset ohjeet tiedostoformaateista kannattaa tarkistaa etukäteen. Tallenteet tulee koko ajan järjestää ja nimetä niin, että niiden luettelointi, merkitseminen ja kuvailu on mahdollisimman selkeää ja yksinkertaista. Myös varmuuskopiointi tulee ottaa huomioon jo kenttätyössä, ja kopiot tulee siirtää palvelinsäilytykseen esimerkiksi arkistoon heti kun olosuhteet sen sallivat. Hyvin järjestetyn aineiston yhteyteen kuuluvat aineiston kuvailutiedot, metatiedot eli metadata. *Metatiedoilla* tarkoitetaan arkistoinnin ja jatkokäytön kannalta oleellisia taustatietoja, joita ovat yleensä ainakin tarkat tiedot aineiston keruuajasta ja -paikasta sekä nauhoittajasta. Metatietoihin on hyvä liittää myös tiedot kustakin nauhoitustilanteesta, siihen osallistuvien ihmisten määrästä ja esimerkiksi iästä, sukupuolesta ja muusta taustasta sekä peruskuvaus puheenaiheesta ja nauhoituksen sisällöstä. Myös tiedot aineiston ja sen osien julkisuudesta, suojaamisesta ja käyttöoikeuksista kuuluvat taustatietoihin. Kieliarkistot auttavat aineistojen julkisuusluokitusten ja käyttölupamenettelyiden määrittämisessä. Jo nauhoitustilanteessa on siis otettava huomioon paitsi nauhoituksen laatu ja tallennusformaatti myös tarvittavien tietojen kokoaminen aineiston kuvaamista ja luokittelua varten. Samalla on huolehdittava tutkimusluvista, joilla sovitaan myös nauhoitukseen osallistuvien puhujien oikeuksista (ks. myös luku 3.2).

Hyvään tutkimusetiikkaan kuuluu paitsi aineistojen vapaa saatavuus myös tutkimukseen osallistuvien ihmisten yksityisyyden suojaaminen. Yhtäältä aineisto tulisi saattaa vapaasti tutkijoiden ja muiden käyttäjien sekä kieliyhteisön käytettäväksi. Toisaalta aineiston koostajalla on suuri tutkimuseettinen vastuu nauhoituksiin ja tutkimustilanteisiin osallistuneiden ihmisten oikeuksista ja yksityisyydensuojasta. Tämä on

kielentutkimuksessa usein tarkoittanut aineistossa esiintyvien tunnistetietojen, kuten nimien ja ammattien, pseudonymisointia julkaisuissa. Aineistoon kuuluvien video- tai audiotallenteiden anonymisoiminen ei kuitenkaan ole mahdollista pelkkiä nimiä muuttamalla: puheääni itsessään on henkilötieto, ja tallenteet koostuvat monesti keskusteluista, joihin voi sisältyä monenlaisia tunnistetietoja, kuten henkilötietoja, juoruja ja mielipiteitä. Siksi tärkeintä onkin, että aineiston käytössä, käsittelyssä ja arkistoinnissa kunnioitetaan tutkimukseen osallistuneiden henkilöiden toiveita ja heidän kanssaan sovittuja asioita.

Tekijänoikeus kustakin aineiston tuottamisen työvaiheesta, esim. sisällöstä tai annotaatiotasosta, on kyseisen osuuden tekijällä. Lisäksi on tarkistettava, sisältääkö kentällä tehty ääni- tai videotallenne materiaalia, jonka tekijänoikeuden voidaan tulkita kuuluvan jollekulle muulle kuin tallennustilanteeseen osallistuneille, esimerkkinä tutkimus- ja haastattelutilanteen yhteydessä tallennettu lauluesitys. Paitsi tekijänoikeuksien tunnustaminen joissakin tapauksissa myös tarve kielenoppaiden työn tunnustamiseen ylipäänsä saattaa olla ristiriidassa yksityisyydensuojan turvaamiseen tähtäävien käytäntöjen kanssa. Jotkut kielenoppaat haluavat nimenomaisesti esiintyä omalla nimellään ja saada julkisen tunnustuksen kielensä hyväksi tekemästään kulttuurityöstä. Kielenoppaiden määrä, heidän toiveensa anonymiteetistä tai nimellä esiintymisestä, aineiston luonne ja siihen mahdollisesti sisältyvät arkaluonteiset, esim. terveydentilaa koskevat aineistoon päätyneet tiedot vaikuttavat siihen, miten tekijyys ja hankkeeseen osallistumisen tunnustaminen ratkaistaan yksityisyydensuojaa kunnioittaen.

Sama aineisto voi parhaimmillaan hyödyttää vuosikymmenten ajan useita erilaisia yleisöjä: tutkijoita, kieliyhteisön jäseniä, kulttuurista kiinnostuneita, suurta yleisöä (ks. Woodbury 2011). Sama aineisto voi palvella monia kielitieteen lähestymistapoja, kuten antropologista lingvistiikkaa, kieliopin kuvausta, kielitypologiaa, fonetiikkaa, diskurssintutkimusta ja sosiolingvistiikkaa. Myös muut, yllättävätkin tieteenalat voivat hyödyntää kielitieteellistä tutkimusta varten kerättyä aineistoa. Ideaalitapauksessa pankkiin tallennettu kieliaineisto kasvaa korkoa: aineiston jatkokäyttäjät lisäävät oman lähestymistapansa kannalta relevantteja annotaatiotasoja korpukseen ja luovuttavat omat merkintänsä

edelleen aineistokokonaisuuden osaksi. Sujuvan aineistojen yhteiskäytön järjestäminen on suuri haaste, jonka ratkaisemiseksi tarvitaan tutkijayhteisön ja arkistolaitosten yhteistyötä (ks. myös Jalava 2011). Arkistoinnin ja pysyväissäilytyksen lisäksi on myös suunniteltava erikseen, millä tavalla aineisto palautetaan kieliyhteisölle mahdollisimman käyttökelpoisessa muodossa. Kieliyhteisön jäsenet eivät välttämättä pysty lukemaan kielitieteellisiä merkintöjä tai käyttämään aineistoa erikoisohjelmistoilla. Hyvä vaihtoehto voi olla raakadatan eli valikoitujen nauhoitusten luovuttaminen kulttuuri-instituuteille ja -toimijoille kenttätyökohteessa sekä mahdollisuuksien ja tarpeiden mukaan painettujen tai digitaalisten tarinakokoelmien ja muiden materiaalien tuottaminen yhteistyössä kieliyhteisön kanssa heidän ymmärtämällään kirjoitustavalla.

## 6. Yhteenveto

Olemme tässä artikkelissa tarkastelleet aiemmin kuvaamattoman tai vähän tutkitun kielen kuvaamista useista näkökulmista. Olemme esitelleet kielten dokumentointiin ja kieliopin kirjoittamiseen tähtäävän kenttätyön vaiheet ja keskeiset aineistonkeruumenetelmät. Kielitieteellinen kenttätyö on aina monivaiheinen prosessi, joka koostuu varsinaista kenttätyömatkaa edeltävistä valmisteluista, työskentelystä kentällä sekä kenttäaineistojen koostamisesta korpukseksi, aineistojen arkistoinnista, julkaisemisesta eri muodoissa ja palauttamisesta kieliyhteisölle. Kieliopin kirjoittamisessa työskentelyä ohjaavat alusta alkaen kohdeyleisön tarpeet; tavoitteena on selkeä ja lukijakunnalle helposti ymmärrettävä kokonaisuus, joka yhtäältä käsittelee järjestelmällisesti kieliopin kuvauksen keskeisiä osa-alueita ja toisaalta tuo esiin tutkittavan kielen erityispiirteet. Kenttätyön pohjalta kirjoitettujen kielioppien aineisto sisältää tasapainoisessa suhteessa toisiinsa vapaata puhetta ja monenlaisiin lavastettuihin kielenkäyttötilanteisiin, käännöstehtäviin ja kieliopillisuusarvioihin perustuvia esimerkkejä. Tämä mahdollistaa sekä tutkittavan kielen kuvaamisen luonnollisessa käyttökontekstissaan että kaikkien

kieliopin osa-alueiden systemaattisen tarkastelun. Toivomme, että tämä lyhyt katsaus tarjoaa lukijalle yleiskuvan kielten dokumentoinnista ja kieliopin kirjoittamisesta ja kannustaa tutustumaan aiheeseen lisää.

**Aiheesta lisää:**


Austin, Peter & Sallabank, Julia (toim.). 2011. *The Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press.

Bowern, Claire. 2008. *Linguistic Fieldwork. A Practical Guide*. New York (NY): Palgrave Macmillan.

Chelliah, Shobhana L. & de Reuse, Willem J. 2011. *Handbook of Descriptive Linguistic Fieldwork*. Dordrecht: Springer.

Meakins, Felicity, Green, Jennifer & Turpin, Myfany. 2018. *Understanding Linguistic Fieldwork*. London: Routledge.

Payne, Thomas E. & Weber, David J. (toim.). 2006. *Perspectives on Grammar Writing*. Benjamins Current Topics 11. Amsterdam: John Benjamins.


## VIITTEET

1 Toki näinkin pieni prosenttiosuus maailman koko väestöstä on suuri joukko, noin 300 miljoonaa ihmistä. Tämän joukon kesken jakautuu 6 700 kieltä. Suurimmalla osalla maailman kielistä on siis keskimäärin korkeintaan muutama kymmenentuhatta puhujaa, todellisuudessa monella niistä vain joitakin kymmeniä.

2 Hyviä esimerkkejä kielioppiluonnoksista löytyy esimerkiksi Lincom Europa -kustantamon julkaisemasta *Languages of the World* -julkaisusarjasta. Sarjassa ilmestyy 50–100 sivun mittaisia kielioppiluonnoksia eri puolilla maailmaa puhutuista kielistä. Suomalaisten tutkijoiden laatimia kielioppiluonnoksia julkaisusarjassa ovat Janhunen (2003) ja Janhunen, Peltomaa, Sandman ja Xiawu Dongzhou (2008).

3 Nenetsin sanaluokkajaon logiikalle muunnettuna suomen kielen lauseke ’musta koira’ kuuluisi muodossa *mustaava koira*, jossa ’musta’ (*mustata*) olisi verbi ja esiintyisi siksi attribuuttiasemassa partisiippimuodossa. Sitä vastoin lauseke ’valkoinen koira’ rakentuisi samalla periaatteella kuin suomessa todellisuudessakin: *valkoinen koira*, koska ’valkoinen’ olisi nomini.

4 Suunnitellut Wallace Chafe (1980), saatavissa Mary S. Erbaughn verkkosivulla http://www.pearstories.org/ (viitattu 23.12.2017).

5 Mercel Mayerin lapsille suunnatut kuvatarinat 1960- ja 1970-luvuilta, esim. *Frog, Where Are You?* (1969).

6 https://www.eva.mpg.de/lingua/tools-at-lingboard/material_scope_fieldwork_project.php.

7 Tätä artikkelia kirjoitettaessa tärkeitä sivustoja ovat TulQuest (http://tulquest.huma-num.fr, viitattu 1.4.2018), jota ylläpitää French Fédération de typologie et universaux linguistiques sekä Max Planck Institute for Evolutionary Anthropologyn entisen kielitieteen laitoksen sivu (https://www.eva.mpg.de/lingua/tools-at-lingboard/tools.php, viitattu 1.4.2018).

8 Eri tutkimussuuntauksissa on erilaisia annotaatiokäytänteitä. Eri tarkoituksiin tarvitaan erilaisia ja eri tarkkuusasteiden merkintätasoja. Joskus aineiston järjestelmällisestä merkinnästä puhutaan myös aineiston koodaamisena.

9 Lyhenteet: 3SG = yksikön 3. persoona, ELA = elatiivi, PTCP = partisiippi, PST = mennyt aika, PASS = passiivi.

10 https://www.eva.mpg.de/lingua/resources/glossing-rules.php.

11 Kehittäjä ja ylläpitäjä: Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, Hollanti. Saatavissa: https://tla.mpi.nl/tools/tla-tools/elan/ (viitattu 1.4.2018.).

12 Kehittäjät ja ylläpitäjät: Paul Boersma & David Weenink. Phonetic Sciences, University of Amsterdam, Hollanti. Saatavissa: https://www.fon.hum.uva.nl/praat/ (viitattu 23.12.2017).

13 CLARIN ERIC https://www.clarin.eu/.

## KIRJALLISUUS


Aineistonhallinnan käsikirja. Tampere: Yhteiskuntatieteellinen tietoarkisto [ylläpitäjä ja tuottaja]. [verkkojulkaisu]. [viitattu 28.2.2019]. Saatavissa: https://www.fsd.tuni.fi/fi/palvelut/aineistonhallinta/.

Austin, Peter & Sallabank, Julia. 2018. Language documentation and language revitalization: Some methodological considerations. Julkaisussa: Hinton, Leanne & Huss, Leena & Roche, Gerald (toim.) *The Routledge Handbook of Language Revitalization*. Abingdon: Routledge, 207–215.

Berghäll, Liisa. 2010. *Mauwake Reference Grammar*. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-6416-6.

Bergqvist, Henrik. 2016. Complex epistemic perspective in Kogi (Arwako). *International Journal of American Linguistics* 82:1, 1–34.

Bouquiaux, Luc & Thomas, Jacqueline M. (toim.). 1992. *Studying and Describing Unwritten Languages*. Dallas (TX): SIL International.

Bowern, Claire. 2008. *Linguistic Fieldwork. A Practical Guide*. New York (NY): Palgrave Macmillan.

Bradley, David & Bradley, Maya. 2002. Introduction. Julkaisussa: Bradley, David & Bradley, Maya (toim.) *Language Endangerment and Language Maintenance: An Active Approach*. London: Curzon Press, xi–xx.
Brown, Penelope. 1993. The role of shape in the acquisition of Tzeltal Mayan locatives. Julkaisussa: Clark, Eve V. (toim.) *The Proceedings of the Twenty-fifth Annual Child Language Research Forum*. Stanford (CA): Center for the Study of Language and Information, 211–220.
Chafe, Wallace (toim). 1980. *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production*. Norwood (NJ): Ablex.
Chelliah, Shobhana L. & de Reuse, Willem J. 2011. *Handbook of Descriptive Linguistic Fieldwork*. Dordrecht: Springer.
Cialone, Claudia. 2017. Navigation talk and culture in Bininj Kunwok. Esitelmä ALT-konferenssissa Canberrassa joulukuussa 2017.
Comrie, Bernard & Smith, Norval. 1977. Lingua descriptive studies: questionnaire. *Lingua* 42, 1–72.
Crystal, David. 2000. *Language Death*. Cambridge: Cambridge University Press.
Dixon, Robert M. W. 2010. *Basic Linguistic Theory. Volume 1: Methodology*. Oxford: Oxford University Press.
Dryer, Matthew S. 2006. Descriptive theories, explanatory theories and Basic Linguistic Theory. Julkaisussa: Ameka, Felix, Dench Alan & Evans, Nicholas (toim.) *Catching Language: The Standing Challenge of Grammar Writing*. Berlin: Mouton de Gruyter, 207–234.
Epps, Patience. 2008. *A Grammar of Hup*. Mouton Grammar Library 43. Berlin: Mouton de Gruyter.
Evans, Nicholas & Dench, Alan. 2006. Introduction: Catching language. Julkaisussa: Ameka, Felix, Dench Alan & Evans, Nicholas (toim.) *Catching Language: The Standing Challenge of Grammar Writing*. Berlin: Mouton de Gruyter, 1–39.
Frajzyngier, Zygmunt. 2002. *A Grammar of Hdi*. Mouton Grammar Library 21. Berlin: Mouton de Gruyter.
Gerstenberger, Ciprian, Partanen, Niko, Rießler, Michael, & Wilbur, Joshua. 2016. Utilizing language technology in the documentation of endangered Uralic languages. *Northern European Journal of Language Technology* 4, 29–47. Saatavissa: https://nejlt.ep.liu.se/article/view/1660.
Gil, David. 2001. Escaping eurocentrism: Fieldwork as a process of unlearning. Julkaisussa: Newman, Paul & Ratliff, Martha (toim.) 2001. *Linguistic Fieldwork*. Cambridge: Cambridge University Press, 102–132.
Grünthal, Riho. 2010. Matkueita tai yksityisiä tutkijoita. Julkaisussa: Kokkonen, Paula & Kurvinen, Anna (toim.) *Kenttäretkistä tutkimustiedoksi*. Uralica Helsingiensia 4. Helsinki: Suomalais-Ugrilainen Seura, 17–56.
Haugen, Einar. 1971. The ecology of language. *The Linguistic Reporter* Supplement 25, 19–26.
Hellwig, Birgit. 2006. Field semantics and grammar-writing: Stimuli-based techniques and the study of locative verbs. Julkaisussa: Ameka, Felix, Dench, Alan & Evans, Nicholas (toim.) *Catching Language: The Standing Challenge of Grammar Writing*. Berlin: Mouton de Gruyter, 321–358.

Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36, 161–195.
Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? Julkaisussa: Gippert, Jost, Himmelmann, Nikolaus P. & Mosel, Ulrike (toim.) *Essentials of Language Documentation*. Trends in Linguistics. Studies and Monographs 178. Berlin: Mouton de Gruyter, 1–30.
Honkasalo, Sami. 2019. *A Grammar of Eastern Geshiza: A Culturally Anchored Description*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-5733-1.
Jalava, Lotta. 2011. Kohti yhteisiä aineistokäytänteitä. *Virittäjä* 115:1, 98–103.
Jalava, Lotta. 2013. "Adjectives" in Tundra Nenets: Properties of property words. *Journal de la Société Finno-Ougrienne* 94, 37–67. Saatavissa: https://www.sgr.fi/susa/94/SUSA94.pdf.
Janhunen, Juha. 2003. *Khamnigan Mongol*. Languages of the World/Materials 173. München: Lincom Europa.
Janhunen, Juha, Peltomaa, Marja, Sandman, Erika & Xiawu Dongzhou. 2008. *Wutun*. Languages of the World/Materials 466. München: Lincom Europa.
Killian, Don. 2015. *Topics in Uduk Phonology and Morphosyntax*. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-1490-7.
Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: Pennsylvania University Press.
Li, Charles N. & Thompson, Sandra A. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, Los Angeles (CA) & London: University of California Press.
Loughnane, Robyn. 2010. *A Grammar of Oksapmin*. Väitöskirja. Melbourne: The University of Melbourne.
Mayer, Mercel. 1969. *Frog, Where Are You?* New York (NY): Dial Books for Young Readers.
Munro, Pamela. 2006. From parts of speech to the grammar. Julkaisussa: Payne, Thomas E. & Weber, David J. (toim.) *Perspectives on Grammar Writing*. Benjamins Current Topics 11. Amsterdam: John Benjamins, 71–111.
Noonan, Michael. 2006. Grammar writing for a grammar-reading audience. Julkaisussa: Thomas E. Payne & Weber, David J. (toim.) *Perspectives on Grammar Writing*. Benjamins Current Topics 11. Amsterdam: John Benjamins, 113–126.
Ojutkangas, Krista, Larjavaara, Meri, Miestamo, Matti & Ylikoski, Jussi. 2009. *Johdatus kielitieteeseen*. Helsinki: WSOY Oppimateriaalit.
Payne, Thomas E. 1997. *Describing Morphosyntax. A Guide for Field Linguists*. Cambridge: Cambridge University Press.
Payne, Thomas E. & Weber, David J. (toim.). 2006. *Perspectives on Grammar Writing*. Benjamins Current Topics 11. Amsterdam: John Benjamins.
Saarinen, Sirkka. 2007. *Fennougristinen kenttätyö*. [verkkoaineisto]. [viitattu 16.10.2017.] Saatavissa: https://www.kotus.fi/nyt/kolumnit_artikkelit_ja_esitelmat/artikkelit/fennougristinen_kenttatyo.
Sandman, Erika. 2013. Vertailu työvälineenä vähän tutkitun kielen kuvauksessa. Julkaisussa: Kolehmainen, Leena, Miestamo, Matti & Nordlund, Taru (toim). *Kielten vertailun metodiikka*. Suomalaisen Kirjallisuuden Seuran Toimituksia 1387. Helsinki: Suomalaisen Kirjallisuuden Seura.

Sandman, Erika. 2016. *A Grammar of Wutun*. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:ISBN 978-951-51-2634-4.
San Roque, Lila, Gawne, Lauren, Hoenigman, Daria, Miller, Julia Colleen, Rumsey, Alan, Spronck, Stef, Carroll, Alice & Evans, Nicholas. 2012. Getting the story straight: Language fieldwork using a narrative problem-solving task. *Language Documentation & Conservation* 6, 135–174.
San Roque, Lila, Floyd, Simeon & Norcliffe, Elisabeth. 2018. Egophoricity. An introduction. Julkaisussa: Floyd, Simeon, Norcliffe, Elisabeth & San Roque, Lila (toim.) *Egophoricity*. Typological Studies in Language 118. Amsterdam: John Benjamins, 1–77.
Shopen, Timothy (toim.). 1985. *Language Typology and Syntactic Description*. (3 nidettä). Cambridge: Cambridge University Press.
Shopen, Timothy (toim.). 2007. *Language Typology and Syntactic Description*. (3 nidettä, 2. p.). Cambridge: Cambridge University Press.
Silva, Wilson & AnderBois, Scott. 2016. Fieldwork Game Play: Masterminding Evidentiality in Desano. *Language Documentation & Conservation* 10, 58–76.
Simons, Gary F. & Fennig, Charles D. (toim.). 2018. *The Ethnologue: Languages of the World*. (21. p.). Dallas (TX): SIL International. [verkkoaineisto]. [viitattu 1.4.2018]. Saatavissa: https://www.ethnologue.com.
Sippola, Eeva. 2012. *Una Gramática Descriptiva del Chabacano de Ternate*. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-7327-4.
Wilde, Christopher P. 2008. *A Sketch of the Phonology and Grammar of Rājbanshi*. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-4465-6.
Woodbury, Anthony C. 2011. Language documentation. Julkaisussa: Austin, Peter & Sallabank, Julia (toim.) *The Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press, 159–186.
Yliniemi, Juha. 2019. *A Descriptive Grammar of Denjongke (Sikkimese Bhutia)*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-5139-1.

# Luku 14 Kielitypologian menetelmät

*Matti Miestamo*
https://orcid.org/0000-0001-9801-9030
*Kaius Sinnemäki*
https://orcid.org/0000-0002-6972-5216

**Mikä?**

Kielitypologia on tutkimusalue, joka tutkii maailman kielten monimuotoisuutta ja siihen vaikuttavia tekijöitä. Se on empiiristä kielentutkimusta, jossa pääasiallisena aineistona toimivat yksittäiskielten kielioppikuvaukset sekä rinnakkaiskorpukset tai kyselyaineistot. Pääasiallisena teoreettisena lähtökohtana on funktionaalinen ja käyttöpohjainen lähestymistapa kieleen. Kielitypologia kehittää menetelmiä erilaisten ihmiskielten vertailuun ja tuottaa uutta perustietoa ihmiskielten luonteesta ja kielten variaatiosta.

**Katso myös:**

Luku 12 Kielikontaktien tutkimus

Luku 13 Kielten dokumentointi ja kieliopin kuvaus

Luku 15 Historiallinen ja vertaileva kielentutkimus

Luku 17 Kieliopin tutkimus ja kielioppiteoriat

# 1. Johdanto

Tässä artikkelissa pureudumme kielitypologian teoriataustaan ja metodologisiin ratkaisuihin.[1] Kielitypologia vertailee maailman kieliä usein hyvinkin laajojen otosten pohjalta ja pyrkii tätä kautta ymmärtämään kielten ominaisuuksia ja kielen olemusta. Se on siis teoreettista kielitiedettä, jonka empiirinen perusta on maailmanlaajuisessa kielten vertailussa. Keskeisiä kysymyksiä kielitypologille ovat: Mikä on maailman kielille yhteistä – universaalia – ja miten kielet eroavat toisistaan; siis mitkä ovat kieltenvälisen variaation rajat? Mikä on yleistä ja mikä harvinaista maailman kielissä? Millaisia yhteyksiä eli korrelaatioita eri kielenpiirteiden välillä voidaan havaita (kielenpiirteen määrittelystä ks. luku 3.4.)? Miten eri rakennetyypit jakautuvat maantieteellisesti? Kuten millä tahansa muulla (kieli)tieteen alalla näihin kaikkiin kysymyksiin lisätään myös soveltuva miksi-kysymys, eli pyritään selittämään ja ymmärtämään tehtyjä kieltenvälisiä havaintoja ja yleistyksiä. Nimitys *typologia*, joka juontaa juurensa tutkimusalan historiaan, korostaa ehkä liiaksikin luokittelun merkitystä – luokittelu on yksi typologisen tutkimuksen vaihe ja osatavoite matkalla ymmärrykseen ihmiskielestä.

Varsinainen kielitypologinen tutkimus on nimenomaan laajoihin kieliotoksiin pohjautuvaa tutkimusta. Väljemmin ymmärrettynä typologista tai typologisesti suuntautunutta tutkimusta voidaan kuitenkin tehdä myös suppeamman kielten vertailun pohjalta tai vaikkapa yksittäistä kieliryhmää tai kieltä typologisesta näkökulmasta tarkastellen. Tässä artikkelissa keskitymme ensisijaisesti laajaan vertailuun perustuvan typologisen tutkimuksen metodeihin. Suomeksi kielitypologian metodologiasta, erityisesti kielten vertailtavuuden problematiikan kannalta, on aiemmin kirjoittanut Miestamo (2013); käsillä olevassa artikkelissa laajennamme näkökulmaa ja omistamme enemmän tilaa tutkimusprosessin muille vaiheille, mm. aineiston keruulle kielten laajassa vertailussa.

Periaatteessa mitä tahansa kielen ilmiötä voidaan lähestyä laajasta vertailevasta näkökulmasta. Käytännössä eniten typologista tutkimusta on tehty kieliopin keskeisistä osa-alueista, kuten kielten morfologisista ominaisuuksista, sanajärjestyksestä, lauseen osallistujien merkinnästä

ja tempus-, aspekti- ja moduskategorioista. Tällaisia ilmiöitä on metodologisesti helpompi lähestyä, koska niistä yleensä löytyy tietoa hyvinkin suuresta määrästä kieliä. Kielten peruskuvaukset käsittelevät useimmiten ainakin tällaisia keskeisiä kieliopin osa-alueita. On siis luonnollista, että kielitypologinen tutkimus on lähtenyt liikkeelle näistä ilmiöistä. Foneettinen ja fonologinen tutkimus on perinteisesti aina ollut kieltenvälisesti orientoitunutta, vaikka sitä onkin harjoitettu jossain määrin erillään varsinaisesta kielitypologisesta tiedeyhteisöstä. Nyttemmin fonologinen typologia on enenevässä määrin integroitumassa typologian muuhun kenttään. Myös leksikaalinen typologia eli sanamerkitysten ja leksikon ominaisuuksien laaja kieltenvälinen vertailu on viime aikoina saanut kasvavaa huomiota typologiassa. Sellaisia kielen ilmiöitä, joiden kieltenvälinen vertailu vaatii typologilta yksittäisten kielten aineiston syvällistä analyysia, on vaikeampi tarkastella laajasta typologisesta näkökulmasta. Niinpä esimerkiksi erilaisten diskurssi-ilmiöiden typologinen tutkimus on vielä lapsenkengissään. Hyvän kuvan kielitypologiassa perinteisesti käsitellyistä aihepiireistä antaa tutustuminen teoksen *The World Atlas of Language Structures* verkkoversioon osoitteessa https://wals.info (Dryer & Haspelmath 2013).

Yleinen kielitiede pyrkii tekemään kaikkia kieliä koskevia yleistyksiä. Laajamittaista kielten vertailua harjoittava kielitypologia on näin yleisen kielitieteen kentän keskiössä. Kaikkia kieliä koskevia yleistyksiä ei voida tehdä yhden tai vain muutaman kielen aineiston pohjalta. Yleisesti maailman kieliin soveltuvia teorioita ja metodeja ei voida kehittää ottamatta huomioon kieltenvälistä variaatiota maailmanlaajuisessa perspektiivissä. Niillekin tutkimusasetelmille, joissa empiirinen perusta ei käsitä laajaa kielten vertailua, on tärkeää ottaa huomioon kielitypologiassa tehdyt havainnot ja eri kielistä löytyvät ilmiöt. Kielitypologia ruokkii näin teorianmuodostusta kielitieteen eri aloilla. Laajemmassa tieteidenvälisessä kontekstissa voidaan todeta, että pyrkiessään kaikkia kieliä koskeviin yleistyksiin yleinen kielitiede – ja erityisesti kielitypologia – on tärkeä osa sitä monimutkaista palapeliä, jota tieteentekijät kokoavat ymmärtääkseen ihmismieltä ja ihmisen toimintaa.

Kielitypologian ja yksittäiskielten tutkimuksen välillä vallitsee riippuvuussuhde ja hedelmällinen vuorovaikutus. Toisaalta kielitypologit

ovat riippuvaisia yksittäiskielten tutkijoiden tuloksista voidakseen ylipäätään vertailla kieltä. Toisaalta yksittäiskielten tutkijat hyötyvät typologisesta tiedosta, joka auttaa ymmärtämään yksittäiskielen ilmiöitä laajemmassa kontekstissa ja antaa näin eväitä analyysien hiomiselle ja aineiston uudelleentulkinnalle. Mitä tahansa kielen ilmiötä yksittäisessä kielessä tai kieliryhmässä tutkiessaan kannattaa aina tutustua vastaaviin ilmiöihin muissa kielissä, ja mikäli ilmiöstä on olemassa typologista tutkimusta, se tarjoaa helpoimman pääsyn tällaiseen kieltenväliseen tietoon. Pyrimme tässä artikkelissa valottamaan typologian teoreettista ja metodologista taustaa, ei pelkästään niille, joiden oma tutkimus on typologista luonteeltaan, vaan myös niille, jotka käyttävät typologian tutkimustuloksia oman yksittäisiä kieliä koskevan tutkimuksensa taustalla. Tutustuminen typologian teoreettis-metodologisiin taustoihin auttaa ymmärtämään typologisen tiedon luonnetta syvällisemmin.

Artikkeli etenee seuraavasti. Luvussa 2 luomme lyhyen katsauksen kielitypologian historiaan ja historialliseen taustaan sekä muutamiin keskeisiin teoreettisiin kysymyksiin. Luvussa 3 tarkastelemme typologisia tutkimusprosesseja: vertailuperustaa, otoskielten valintaa, luokittelua, jakaumien tarkastelua, kielenpiirteiden välisiä korrelaatioita ja tulosten selittämistä. Luvussa 4 pohdimme lyhyesti typologian paikkaa kielitieteen kentällä suhteessa lähialoihin sekä teemme lyhyen yhteenvedon artikkelin teemoista.

## 2. Kielitypologian historiaa ja teoreettista taustaa

### 2.1. Katsaus kielitypologian historiaan

Kielten systemaattinen ja laajamittainen vertailu on lähestymistapana suhteellisen nuori. Se alkoi varsinaisesti Joseph Greenbergin tutkimuksista 1960-luvulla. Kaikkien kielten yhteyttä etsivä universalistinen ajattelu on sen sijaan paljon vanhempaa perua eurooppalaisessa

kontekstissa. Se juontaa juurensa jo antiikin ajattelijoihin, ja omat universaalikielioppimallinsa ovat myöhemmin esittäneet mm. modistien nimellä tunnettu koulukunta myöhäiskeskiajalla sekä Port Royal -koulukunta, joka vaikutti 1600-luvulla Ranskassa (ks. Itkonen 1991). Yhteistä näille historiallisille suuntauksille on, että universaaleista päämääristä huolimatta tarkastelu on rajoittunut lähinnä kulloiseenkin valtakieleen. Tässä mielessä ne vertautuvat suoraan chomskylaiseen generatiiviseen kielioppiin, joka nykysuuntauksista ehkä selvimmin perustaa olettamuksensa kielen universaaleista ominaisuuksista yhden (tai korkeintaan muutaman) kielen analyysille. Kielitypologian lähtökohta on täysin vastakkainen: tutkittaessa maailman kielten erilaisuutta ja samuutta on empiirisen aineiston katettava maailman kielten kirjo mahdollisimman laaja-alaisesti ja tasaisesti.

Kielten vertailusta alettiin kiinnostua enenevässä määrin 1800-luvulle tultaessa. Keskeisimpiä huomion kohteita oli kielten morfologinen rakenne. Klassinen morfologinen typologia, jonka näkyvimpiä edustajia olivat mm. August (1767–1845) ja Wilhelm (1772–1829) von Schlegel sekä Wilhelm von Humboldt (1767–1835), jakoi kielet kolmeen perustyyppiin: isoloiviin kieliin, joissa ei ole sidonnaista morfologiaa lainkaan eli kukin sana koostuu yhdestä morfeemista, sekä agglutinoiviin ja fusionaalisiin kieliin, joissa molemmissa sanat sisältävät affikseja, mutta edellisissä vartaloiden ja affiksien muodoissa ei ole vaihtelua, kun taas jälkimmäisissä esiintyy erilaisia morfofonologisia ilmiöitä. Myöhemmin näihin kolmeen perustyyppiin lisättiin vielä neljäs eli polysynteettiset kielet, joissa affiksien määrä sanassa on huomattavan suuri ja yksi sana voi äärimmillään vastata pitkääkin lausetta kolmen muun tyypin kielissä. Huomattakoon, että kun puhuttiin kielityypeistä eikä rakennetyypeistä, ajatuksena oli, että kielen morfologinen tyyppi luonnehtisi kieltä yleisellä tasolla. Tuon ajan typologiseen ajatteluun liittyi myös etnopsykologinen ote, jossa kielen tyypin ajateltiin heijastelevan puhujayhteisön kulttuurista kehitysastetta ja ajattelutapaa. 1800-luvun lopulla ja 1900-luvun alkupuoliskolla kiinnostus typologiseen vertailuun oli vähäisempää; 1800-luvulla dominoi historiallis-vertaileva kielitiede, ja 1900-luvulla siirryttiin vähitellen strukturalismiin. Typologista ajattelua veivät kuitenkin strukturalismin rinnalla eteenpäin esimerkiksi

Edward Sapir (1884–1939), joka kehitti mm. morfologisen typologian luokitteluita analyyttisempaan suuntaan tunnistamalla erillisiä parametrejä aiemmin yhtenäisempinä esitettyjen morfologisten tyyppien takana (ks. Sapir 1921), sekä Roman Jakobson (1896–1982) ja muutamat muut Prahan koulukunnan edustajat.

Kiinnostus kielten vertailuun ja kielten universaaleihin ominaisuuksiin heräsi toden teolla 1960-luvulla, kun Joseph Greenberg (1915–2001) toi laajan systemaattisen kielten vertailun metodit kielitieteeseen. Eräänlaisena lähtölaukauksena laajempaan tietoisuuteen voidaan pitää Dobbs Ferryssä vuonna 1961 järjestettyä konferenssia *Universals of Language*, johon Greenberg oli kutsunut useita tunnettuja kielitieteilijöitä ja lähialojen edustajia. Tuon konferenssin pohjalta toimitetussa kirjassa ilmestyi Greenbergin (1963) artikkeli ”Some universals of grammar with particular reference to the order of meaningful elements”, joka loi pohjan systemaattiseen kielen vertailuun perustuvalle typologiselle metodologialle ja greenbergiläiselle universaalikäsitykselle. Vuosina 1967–1976 Greenberg johti Stanfordissa Universals of Human Language -projektia, jonka tulokset on vedetty yhteen neliosaisessa kirjasarjassa *Universals of Human Language* (Greenberg 1978). Näiden tapahtumien myötä moderni kielitypologia oli saanut alkunsa (Greenbergistä ks. tarkemmin Miestamo 2010).

Myöhemmistä kielitypologian kehityksen virstanpylväistä mainittakoon seuraavat. 1990-luvulla EU:n rahoittaman EUROTYP-projektin tuloksena syntyi kokoelma kirjoja, joissa kartoitettiin laajasti Euroopan kielten ominaisuuksia kieliopin eri osa-alueilla. Euroopan kieliä tarkasteltiin maailmanlaajuisessa typologisessa perspektiivissä, joten samalla tehtiin laajempaakin typologista tutkimusta. Vuonna 1994 perustettu Association for Linguistic Typology aloitti säännöllisen konferenssisarjansa vuonna 1995 ja alkoi julkaista *Linguistic Typology* -lehteä vuonna 1997. Leipzigin evoluutioantropologian Max Planck -instituutissa vuonna 1997 aloittanut kielitieteen laitos muodostui merkittäväksi kielitypologisen tutkimuksen keskukseksi – Max Planckin politiikan mukaisesti laitos lakkautettiin vuonna 2015 sen johtajan Bernard Comrien jäädessä eläkkeelle. Vuonna 2005 ilmestyi edellä mainittu, noin 40 kirjoittajan yhteistyönä maailman kielten ominaisuuksia kartoittava *The World Atlas*

*of Language Structures* (WALS) (Haspelmath, Dryer, Gil & Comrie 2005). WALS on näkyvä esimerkki siitä viimeaikaisesta kehityksestä, jossa kielten ja kielenpiirteiden maantieteelliseen levinneisyyteen liittyvät kysymykset ovat enenevässä määrin nousseet kielitypologian keskiöön. Moderni kielitypologia syntyi 1960-luvulla vastaamaan kysymykseen kieltenvälisen variaation rajoista ja kielten universaaleista eli siitä, millaiset kielet ovat mahdollisia ja millaiset eivät ja miten tätä kieltenvälistä variaatiota voidaan selittää (etupäässä) funktionaalisesti. Vuosituhannen vaihteen jälkeen on yleistynyt Johanna Nicholsin (1992) klassikkoteoksesta lähtölaukauksensa saanut populaatiotieteellinen näkökulma, jossa rakennetyyppien levinneisyyttä tarkastellaan populaatioina. Bickelin (2007) sanoin kysymys ”mikä on kielessä mahdollista?” on saanut seurakseen kysymykset ”mitä, missä ja miksi?”. Siinä missä ennen otannassa pyrittiin eliminoimaan kielisukulaisuuteen ja kielikontakteihin perustuvat yhtäläisyydet, jotta päästäisiin käsiksi kielen perimmäisiin ominaisuuksiin, nykyisin areaalinen ja genealoginen näkökulma ovat itsessään usein mukana typologisessa tutkimuksessa, ja tyyppien yleisyyttä ja levinneisyyttä halutaan ymmärtää niiden taustalla olevien historiallisten prosessien kautta (ks. tarkemmin Bickel 2007; ks. Kielikontaktien tutkimus tässä kirjassa). Palaamme näihin eri suuntauksiin luvussa 3. Kielitypologian historiasta voi lukea lisää esimerkiksi Ramatin (2010) ja Graffin (2010) katsausartikkeleista.

Suomalaisen kielitieteen kontekstissa voidaan typologisesti orientoituneen tutkimuksen varhaisia juuria nähdä jo 1800-luvulla Anders Sjögrenin ja Matthias A. Castrénin itään suuntautuneissa tutkimusmatkoissa (ks. myös Antropologinen lingvistiikka tk.). Ne aloittivat kansainvälisesti uraauurtavan etnolingvistisen tradition, jossa tarkasteltiin ja vertailtiin eurooppalaisista kielistä typologisesti monella tavalla poikkeavia kieliä ja kulttuureita. Tuloksena syntyi valtavasti typologisesti kiinnostavaa aineistoa pohjoisen Euraasian kielistä. Vertailu oli kuitenkin päämääriltään ensisijaisesti historiallis-vertailevaa, ei typologista. Fennougristiikassa ja fennistiikassa diakroninen näkökulma dominoi aina 1960-luvulle asti. Modernin typologian aikakaudella typologinen näkökulma alkoi rantautua Suomeen. Mikko Korhosen tuotannossa kielitypologinen näkökulma oli vahvasti läsnä (ks. Salminen 1996).

Yllämainittuun EUROTYP-hankkeeseen osallistui suomalaisista kielitieteilijöistä mm. Marja Leinonen, Jouko Lindstedt, Hannu Tommola ja Maria Vilkuna. Turun yliopiston yleisen kielitieteen oppiaineessa Esa Itkonen ja Anneli Pajunen toivat typologista näkökulmaa vahvasti esille (ks. Pajunen 1998; 2000; Itkonen 1997). Pirkko Suihkonen pohti väitöskirjassaan (1990) korpustutkimuksen mahdollisuuksia kielitypologiassa. Ensimmäisiä varsinaisia laajoihin otoksiin perustuvia typologisia tutkimuksia olivat Seppo Kittilän transitiivisuutta (2002) käsittelevä ja Matti Miestamon negaatiota (2000; 2003) käsittelevä tutkimus. Kvantitatiivista näkökulmaa suomalaiseen kielitypologiaan on tuonut erityisesti Kaius Sinnemäki (2008; 2011). Annu Marttilan (2010) leksikaalista ikonisuutta lintujen nimissä tarkastelevaa laajaa vertailevaa tutkimusta voidaan pitää leksikaalisen typologian alaan kuuluvana. Myös Ksenia Shagalin (2019) tuore tutkimus partisiippien typologiasta on herättänyt kansainvälistä huomiota.

## 2.2. Teoreettista taustaa

Ylempänä typologiaa on luonnehdittu maailmanlaajuiseksi vertailevaksi kielitieteeksi ja teoreettiseksi kielitieteeksi, jonka empiirinen perusta on maailmanlaajuisessa kielten vertailussa. Nämä luonnehdinnat sinällään eivät vielä kerro paljoakaan siitä, millaisia teoreettisia taustaoletuksia typologisella tutkimuksella on. Toki empiirinen metodologinen lähestymistapa asemoi typologian rationalismi-empirismijaottelussa empirismin puolelle. Periaatteessa laajaan vertailuun perustuvaa typologista tutkimusta voidaan tehdä hyvinkin erilaisista teoreettisista lähtökohdista. Tavallisimmin kuitenkin typologista tutkimusta harjoitetaan funktionalistisessa viitekehyksessä. Myöskään formaaleja metakieliä ei yleensä käytetä, vaan analyysit esitetään mahdollisimman laajasti ymmärretyn ja jaetun käsitteistön puitteissa. Typologinen vertailu on luonteeltaan synkronista, mutta diakroninen näkökulma on vahvasti läsnä typologisessa tutkimuksessa, etenkin typologisten havaintojen selittämisessä. Keskeinen teoreettinen kysymys on myös yksittäiskielten ja universaalin suhde ja kielellisten kategorioiden luonne kielikohtaisina vs. universaaleina

sekä edelleen se, miten erilaiset kategoriakäsitykset vaikuttavat kielen vertailtavuuteen konkreettisen metodin tasolla. Tässä luvussa luomme lyhyen katsauksen kielitypologian keskeiseen teoreettiseen taustaan.

Kun eri lähestymistapoja kieleen ja kielioppiin luonnehditaan teoreettisen lähestymistavan mukaan joko formalistisiksi tai funktionalistisiksi, typologia identifioituu yleensä selvästi jälkimmäiseen ryhmään (ks. Johdannon luku 3 tk.). Kuten yllä todettiin, laajaa vertailua voidaan toki tehdä monenlaisissa teoriakehyksissä, ja esimerkiksi generatiivinen kielitiede on viime aikoina alkanut kiinnittää enemmän huomiota kieltenväliseen evidenssiin. Laajat vertailevat tutkimukset tällaisista formalistisista lähtökohdista ovat kuitenkin harvinaisia. Yleensä typologien keskeisenä teoreettisena lähtökohtana on, että kielen funktio, kielenkäyttö, vaikuttaa kielen rakenteeseen, ja näin selityksiä kieltenvälisille yleistyksille etsitään ensisijaisesti funktion tasolta eli semantiikasta, pragmatiikasta ja kognitiivisista prosesseista. Jo Greenbergillä (1966) oli vahvasti käyttöpohjainen näkemys kielestä, mistä selkeimpänä esimerkkinä on se, että hän nosti erityisesti ilmiöiden frekvenssin tärkeäksi selittäväksi tekijäksi. Typologian selitysperusteisiin palaamme luvussa 3.7.

Kielten laajahkoa vertailua harjoittavat myös historiallis-vertailevan kielitieteen edustajat. Typologinen vertailu eroaa tästä tavoitteidensa osalta: typologiassa vertailun päämäärä ei ole päästä kiinni kielten historialliseen kehitykseen ja sukulaisuussuhteisiin vaan kuvata kieltenvälistä variaatiota ja ymmärtää ihmiskielen perusperiaatteita laajan vertailun pohjalta. Typologia on siis ensisijaisesti synkronista, siinä missä historiallis-vertailevan kielitieteen tavoitteet ovat lähtökohtaisesti diakroniset. Diakronialla on kuitenkin keskeinen rooli typologiassa. Jo Greenberg (1978) puhui dynaamisesta typologiasta ja tarkasteli mm. sitä, millaiset muutokset rakennetyypistä toiseen ovat mahdollisia, ja tällaiset kysymyksenasettelut ovat tulleet entistä tärkeämmiksi makrotason synkronisten jakaumien selittäjinä. Havaittujen rakennetyyppien jakaumien selittämisessä on oleellista katsoa niiden historiallisen kehityksen yksityiskohtia ja tyypillisiä kieliopillistumiskehityksiä. Vaikka itse vertailu onkin yleensä synkronista, typologisissa tutkimuksissa on usein vahva diakroninen komponentti mukana. Kielen funktionaaliset aspektit vaikuttavat kielen muotoon ja rakenteeseen historiallisten prosessien

kautta, ja typologien selitysmalleissa yhdistyvät näin historialliset ja funktionaaliset selitykset.

Typologiassa, kuten myöskään deskriptiivisessä kielitieteessä (ks. Kielten dokumentointi ja kieliopin kuvaus tk.), ei yleensä käytetä teoriaspesifejä, formalisoituja metakieliä. Sen sijaan analyysit esitetään käyttäen sellaista terminologiaa, joka on kielitieteilijöiden keskuudessa mahdollisimman laajalti jaettua ja ymmärrettyä. Dixon (1997; 2009) ja Dryer (2006) ovat esittäneet, että tällaista yhteistä teoriapohjaa voidaan kutsua *peruskielioppiteoriaksi* (engl. *Basic Linguistic Theory*, BLT). Peruskielioppiteorialla on juurensa perinteisessä kieliopissa, jonka päälle se on eri aikakausien kielitieteellisen tutkimuksen tuloksia hyödyntäen rakentanut modernin, eri kielten ilmiöiden analyysiin ja kuvaukseen laajalti soveltuvan käsitteistön ja terminologian. Peruskielioppiteoria on yhteinen, kumulatiivinen teoriapohja, jonka osaksi pääsee kulloinkin vallalla olevista teorioista käsitteitä, jotka osoittautuvat kestäviksi ja hyödyllisiksi; jotkin käsitteet taas elävät vain lyhyen aikaa eivätkä päädy osaksi peruskielioppiteoriaa. Nykyisin peruskielioppiteoria mainitaan usein eksplisiittisesti teoriataustaksi typologisissa tutkimuksissa ja deskriptiivisissä kieliopeissa.

Keskustelussa typologisen vertailun teoreettisista lähtökohdista on viime aikoina noussut keskeisenä teemana esiin pohdinta kielellisten kategorioiden luonteesta: ovatko kielelliset kategoriat kielille yhteisiä vai kielikohtaisia. **Universalistinen** kategoriakäsitys olettaa kaikille kielille yhteisten kategorioiden olemassaolon, kun taas **partikularistinen** käsitys pitää kielellisiä kategorioita kielikohtaisina. Edellisen mukaan siis sellaiset kategoriat kuin vaikkapa adjektiivi, subjekti tai genetiivi olisivat kielille yhteisiä ja kunkin yksittäisen kielen kategoriat olisivat näiden yhteisten kategorioiden kielikohtaisia reaalistumia. Jälkimmäinen käsitys taas olettaa, että kategorioita ei kieltenvälisessä merkityksessä ole olemassa, vaan ne ovat aina kielikohtaisia. Universalistinen kategoriakäsitys on vallalla nykykielitieteessä varsinkin erilaisten formalististen kielioppimallien kannattajien parissa. Niistä erityisesti mainittakoon chomskylainen generativismi, jossa kategoriat ovat synnynnäisen universaalikieliopin määräämiä. Myös esimerkiksi perinteinen, antiikin kielitieteestä juurensa juontava ja 1800-luvulle asti yleinen tapa kuvata eri kieliä latinan mallin mukaisesti voidaan nähdä universalistisena

lähestymistapana. Partikularistinen kategoriakäsitys on ollut erityisen näkyvä 1900-luvun alkupuolen amerikkalaisessa antropologisessa kielitieteessä ja strukturalismissa, jossa tutkijat saivat tarkasteltavakseen valtavan kirjon eurooppalaisista kielistä radikaalistikin poikkeavia kieliä; reaktio tähän kielten moninaisuuteen oli Martin Joosin (1957, 96) sanoin näkemys, että "kielet voivat erota toisistaan rajattomasti ja ennustamattomasti".[2] Chomskylaisen generativismin nousu käänsi valtavirta-ajattelua vahvasti universalistiseen suuntaan, mutta viime vuosikymmeninä funktionalistisessa kielitieteessä ja etenkin typologiassa on otettu jälleen eksplisiittisesti partikularistisia kantoja (ks. esim. Lazard 1992; Dryer 1997; Haspelmath 2007). Funktionaalis-typologinen partikularistinen näkemys ei kuitenkaan oleta aiempien amerikkalaisten strukturalistien tavoin, että kielten välinen variaatio olisi satunnaista. Kategoriat ovat kielikohtaisia, mutta niiden välillä on yhtäläisyyksiä ja samankaltaisuutta, jonka taustalla on kielille yhteinen funktionaalinen taso – merkitys, käyttö ja kommunikaatiotarpeet. Tutkijan käsitys kategorioiden luonteesta vaikuttaa suoraan vertailun metodologisiin perusteisiin, joihin palataan seuraavassa luvussa.

# 3. Typologinen tutkimusprosessi

Tässä luvussa tarkastelemme lähemmin typologisen tutkimusprosessin eri vaiheita. Funktionalistis-typologinen tutkimusprosessi voidaan tyypillisesti tiivistää alla olevaan seitsemään eri vaiheeseen (soveltaen Stassen 1985, 1–23; Miestamo 2013). Tässä esitettyjä vaiheita edeltää luonnollisesti tutkimuksen tavoitteiden ja tutkimuskysymysten määrittely (suluissa luku, jossa käsittelemme kyseistä tutkimusprosessin vaihetta).

1) Typologisen tutkimusprosessin vaiheet
   I. Aineiston koonti
      1. Määritellään tutkimuksen kohde (3.1.)
      2. Valitaan kieliotos (3.2.)

3. Hankitaan tutkimusmateriaali (3.3.)

II. Aineiston analyysi ja yleistysten tekeminen

1. Analysoidaan ja luokitellaan materiaali (3.4.)
2. Kuvaillaan rakennetyyppien yleisyys ja maantieteellinen jakauma (3.5.)
3. Selvitetään tilastolliset korrelaatiot (3.6.)

III. Tulosten selittäminen (3.7.)

Tämä representaatio on toki yksinkertaistus käytännön työstä. Tutkimus ei useinkaan etene näin suoraviivaisesti, vaan eri vaiheita voidaan työstää samanaikaisesti tai aiempiin vaiheisiin voidaan joutua palaamaan useamman kerran. Analyysi- ja luokitteluvaiheessa saatetaan palata materiaalin keräämiseen, jos alkuperäinen lähde ei ollut riittävä esiin nousseiden kysymysten ratkaisemiseksi. Jos analyysivaiheessa lähdemateriaali osoittautuu muutoinkin riittämättömäksi, typologi voi joutua vaihtamaan otoskielen ja täten palaamaan otantavaiheeseen. Joskus analyysivaiheessa täytyy myös tarkentaa itse tutkimuskohteen määrittelyä, jos se ei ole osoittautunut riittävän tarkaksi. Seuraavassa kuvaamme tutkimusprosessia vaihe vaiheelta.

## 3.1. Tutkimuskohteen määrittely

Typologinen tutkimus lähtee liikkeelle tutkimuskohteen määrittelystä. Oleellinen kysymys on, miten tunnistaa tutkimuksen kohteena oleva ilmiö hyvinkin erilaisissa kielissä, jotta vertailu kohdistuu samaan ilmiöön vertailtavissa kielissä. Tätä ongelmaa kutsutaan typologisessa kirjallisuudessa *kieltenvälisen identifikaation ongelmaksi* (ks. esim. Stassen 2010). Edellisessä luvussa totesimme, että typologisen tutkimuksen yleinen teoreettinen taustaoletus on, että kielelliset kategoriat ovat kielikohtaisia, siis kategoriakäsitys on partikularistinen. Kun kielille yhteisiä kategorioita ei ole olemassa, ei vertailtavan tutkimuskohteen määrittelyä voida perustaa sellaisille.

Jo varhain todettiin, että kielellisen vertailun perustaminen puhtaasti muodollisiin kategorioihin on ongelmallista (Greenberg 1963; Stassen

1985). Mikäli halutaan vertailla eri kielten adjektiiveja, ei niitä voida tunnistaa sellaisten yhdestä tai muutamasta kielestä peräisin olevien muodollisten kriteerien avulla kuin että niillä on morfologisesti ilmaistut vertailuasteet positiivi (*hyvä*), komparatiivi (*parempi*) ja superlatiivi (*paras*), että ne tarvitsevat predikaatteina toimiessaan *olla*-verbin (esim. *Erkki on pitkä*) ja että ne voivat esiintyä intensifikaattorien, kuten *erittäin* ja *sangen*, kanssa. Ominaisuussanojen ilmaisussa on paljon kieltenvälistä variaatiota, ja tällainen muotoon perustuva määritelmä jättää ulkopuolelleen esimerkiksi ne lukuisat kielet, joissa ominaisuussanojen vertailu toteutetaan muuten kuin morfologisesti tai joissa ominaisuussanat eivät ole nomineja vaan toimivat syntaktisesti kuten verbit. Jo varhain siis havaittiin, että kieltenvälisen vertailun on tavallisesti perustuttava muodon sijasta funktioon (ks. Stassen 2010; Miestamo 2013).

Tätä metodologista käytäntöä on viime aikoina kirkastanut Haspelmathin (2010) eksplikoima kuvauskategorioiden (engl. *descriptive categories*) ja vertailukäsitteiden (engl. *comparative concepts*) välinen eronteko. Kuvauskategorioiden kohteena ovat yksittäisten kielten kielikohtaiset kategoriat, joiden semanttiset ja morfosyntaktiset ominaisuudet ovat kullekin kielelle uniikkeja. Nämä ovat osa kunkin kielen systeemiä, ja ne ovat psykologisesti ja sosiaalisesti tosia siinä mielessä, että ne ovat olemassa kunkin kielen puhujayhteisön jakamina konventioina. Kielentutkijan esittämien kuvauskategorioiden psykologinen ja sosiaalinen totuus puolestaan riippuu siitä, miten hyvin tutkija on onnistunut tiivistämään todellisuuden esittämiinsä kategorioihin. Vertailukäsitteet sen sijaan ovat tutkijan metodologisia työkaluja, joiden tavoitteena on toimia kieltenvälisten yleistysten välineenä, eikä niillä ole suoraa relevanssia minkään yksittäisen kielen kuvaukselle. Epäsuorasti ne toki voivat olla hyvinkin relevantteja – kuten alussa totesimme, typologisesta tiedosta on paljon hyötyä kielenkuvauksen prosessissa (ks. myös luku 4). Siinä missä tutkijan esittämät kuvauskategoriat voivat olla oikeita tai vääriä suhteessa kunkin kielen todellisiin kielikohtaisiin kategorioihin, vertailukäsitteet eivät ole samanlaisessa suhteessa todellisuuden kanssa. Metodologisina työkaluina niitä arvioidaan ensisijaisesti sen mukaan, miten hyvin ne soveltuvat kulloisenkin tutkimuskysymyksen selvittelyyn, esimerkiksi miten selkeästi ja luotettavasti vertailukäsite rajaa tutkittavan ilmiön

hyvinkin erilaisista otoskielistä ja missä määrin sen määritelmään perustuva tutkimus tuottaa teoreettisesti kiinnostavia tuloksia.

Kielten kategoriat ovat kielikohtaisia, mutta niiden välillä voidaan havaita paljonkin samankaltaisuutta. Funktionalistisessa teoriakehyksessä kielten kategorioiden samankaltaisuuden nähdään perustuvan kielten keskenään jakamiin funktionaalisiin ominaisuuksiin. Kieltenvälinen vertailu pohjaa näin ollen ensisijaisesti universaalin tason semanttisiin ja pragmaattisiin funktioihin, joita voidaan lisäksi tarkentaa muotoon liittyvillä kriteereillä. Haspelmathin (2010, 670) mukaan vertailukäsitteiden määrittelyssä voidaan käyttää kolmenlaisia käsitetyyppejä:

- konseptuaalis-semanttiset käsitteet (esim. *ominaisuus* alla olevan laatikon esimerkkimääritelmässä)
- yleiset muodolliset käsitteet (esim. *edeltävä* määriteltäessä esim. prepositiota)
- muut vertailukäsitteet (esim. *lekseemi* alla olevan laatikon esimerkkimääritelmässä).

Laatikon esimerkki havainnollistaa vertailukäsitteen määrittelyä.

**Esimerkki:**

Haspelmath (2010, 670) määrittelee vertailukäsitteen *adjektiivi* seuraavasti:

> Adjektiivi on lekseemi, joka merkitsee deskriptiivistä ominaisuutta ja jota voidaan käyttää rajoittamaan substantiivin tarkoitetta.

Määritelmässä esiintyvät vertailukäsitteet *lekseemi* ja *substantiivi* sekä konseptuaalis-semanttiset käsitteet *deskriptiivinen ominaisuus* ja *rajoittaa tarkoitetta*. Tämä määritelmä rajaa ulkopuolelle esimerkiksi sanat *kauneus* ja *koko*, jotka ovat kyllä lekseemejä ja viittaavat ominaisuuteen mutta joita ei käytetä substantiivin määritteenä rajaamaan sen tarkoitetta. Määritelmä rajaa ulkopuolelle myös konstruktion *joka oli kaunis*, joka kyllä viittaa ominaisuuteen ja rajaa substantiivin tarkoitetta mutta joka ei ole lekseemi vaan relatiivilause.

Kieltenvälisen identifikaation ongelma ja kielten vertailtavuus ovat keskeisimpiä metodologisia kysymyksiä kielitypologiassa. Emme kuitenkaan kirjoita niistä tässä tämän pidemmin, koska Miestamon (2013) suomenkielinen esitys aiheesta on avannut problematiikkaa konkreettisesti ja runsain esimerkein.

## 3.2. Otanta

Kun tutkimuskohde on määritelty, seuraava vaihe tutkimusprosessissa on valita kielet, joista tutkimusmateriaali kerätään ja analysoidaan. Ei ole kuitenkaan itsestään selvää, mitkä maailman noin 7 000 kielestä pitäisi valita tutkimukseen. Kielten systemaattista valintaa ja siihen liittyviä periaatteita kutsutaan *otannaksi*, ja se on laajimmin käsitelty aihe typologisessa metodologiassa (ks. esim. Cysouw 2005, 554–555). Joskus eksploratiivisessa eli aiemmin heikosti tunnetun tai tuntemattoman ilmiön kartoittavassa tutkimuksessa käytetään ns. mukavuusotantaa, eli kieliä otetaan mukaan tutkimukseen ilman systemaattisia otantaperiaatteita, mutta tämä ei ole nähdäksemme suositeltavaa edes tutkimuksen pilottivaiheessa. Otannassa kielten valintaa siis rajoitetaan etukäteen erilaisin periaattein, jotta saadun otoksen avulla voitaisiin tehdä yleistyksiä.

Typologiassa käytetyt otantamenetelmät voidaan jakaa karkeasti kahteen luokkaan niiden tarkoituksen mukaan: variaatio-otantaan ja todennäköisyysotantaan. Variaatio-otantaa käytetään, kun pyritään kartoittamaan jonkin kielenpiirteen kieltenvälinen variaatio. Tällöin on tavoitteena saada mahdollisimman kattava käsitys vertailukäsitteen avulla määritellyn piirteen monimuotoisuudesta ja myös sen kaikkein harvinaisimmista ilmenemistyypeistä. Variaatio-otanta hyödyttää etenkin kielioppiteorian validoimista: pätevän kielioppiteorian on kyettävä selittämään myös kaikkein harvinaisimmat kielenpiirteiden tyypit, ja ne löytyvät parhaiten variaatio-otannan avulla (Miestamo, Bakker & Arppe 2016, 233–234). Todennäköisyysotantaa puolestaan käytetään universaalien tutkimuksessa, kun halutaan testata kielenpiirteen jakaumia tilastollisesti. Tällöin tavallinen lähestymistapa otantaan on satunnaisotanta. Klassiset tilastomenetelmät vaativat, että tutkimusyksiköillä, joka

typologiassa tarkoittaa yleensä kieliä, on sama todennäköisyys tulla valituksi otokseen; tähän tavoitteeseen päästään satunnaisotannan avulla. Lisäksi tutkimusyksiköiden pitäisi olla toisistaan riippumattomia. Toisista kielistä riippuvat piirteet sotkevat tilastollisten päätelmien tekemistä, joten lähisukukielet ja lähinaapurit tulisi sulkea otoksen ulkopuolelle. Pohjimmiltaan otantamenetelmien valintaa ohjaa siis tutkimuksen tavoite.

Riippumatta otantamenetelmästä typologisen otannan pääperiaatteena on, että otos edustaisi perusjoukkoa (esim. tällä hetkellä puhuttavia maailman kieliä) ja sen monimuotoisuutta mahdollisimman tasaisesti ja kattavasti. Tyypillisesti kattavaan ja tasaiseen edustukseen pyritään käyttämällä ositettua otantaa (engl. *stratified sampling*), jonka avulla tunnettuja vinoumia kielenpiirteiden jakaumissa pystytään rajoittamaan. Ositetussa otannassa kielet jaetaan ositteisiin, esim. kieliperheisiin ja maantieteellisiin alueisiin, joista kielet poimitaan otokseen – todennäköisyysotannassa yleensä satunnaisotannalla. Ositus voidaan tehdä joko yhdellä tai useammalla tasolla, yleensä ainakin genealogisesti eli kielisukulaisuuden perusteella. Tällöin kielet jaetaan kielikuntiin jonkin genealogisen luokittelun mukaisesti ja kielet valitaan otokseen esimerkiksi kielikunnan sisäisen rakenteen perusteella (ks. Rijkhoff & Bakker 1998) tai valitsemalla yksi kieli jokaisesta genuksesta. Genukset muodostavat maailmanlaajuisesti vertailukelpoiseksi tarkoitetun luokittelun tason, joka vastaa karkeasti indoeurooppalaisen kielikunnan päähaaroja, kuten germaanisia tai romaanisia kieliä (Dryer 1992). Toinen ositustaso on yleensä alueellinen. Tällöin kielet jaetaan samankokoisiin maantieteellisiin alueisiin, kuten Afrikka, Euraasia, Oseania, Australia, Pohjois-Amerikka ja Etelä-Amerikka, ja näiltä alueilta valitaan kieliä esimerkiksi suhteessa kielten genealogiseen monimuotoisuuteen kullakin alueella (Miestamo 2005).

Ositettu otanta on tyypillinen lähestymistapa otantaan kaikissa ihmistieteissä, mutta sillä on myös ongelmansa. Typologiassa keskeinen ongelma liittyy genealogiseen ositukseen: ei ole itsestään selvää, mitä genealogista luokittelua tulisi käyttää, ja otannan tulos voi riippua käytetystä luokittelusta. Toinen ositetun otannan ongelma liittyy synkroniseen ja dynaamiseen typologiaan, joita käsittelemme lyhyesti alla.

Otoksen muodostamisessa keskeinen kysymys on myös otoksen koko. Kun aineistoa analysoidaan sadoista kielistä, yhdestä kielestä ei voi analysoida kuin kapean alueen tai työmäärä kasvaa hallitsemattomaksi. Jos taas otoskoko liikkuu maksimissaan muutamassa kymmenessä kielessä, tutkimuskohteen pystyy analysoimaan syvällisemmin. Kysymyksenasettelu ja otoskoko rajaavat näin ollen omalta osaltaan typologista tutkimusta. Sopivan otoskoon löytäminen on tärkeää, koska liian pieni otos ei riitä luotettaviin yleistyksiin ja liian suuri otos taas hukkaa arvokasta tutkimusaikaa. Perkins (1989) esittää, että yleensä 100 kieltä on riittävä otoskoko useimpiin typologisiin tutkimuksiin. Asia ei kuitenkaan ole niin yksinkertainen, sillä riittävä otoskoko riippuu rakennetyyppien määrästä typologiassa: mitä enemmän rakennetyyppejä kielenpiirteen typologisessa luokittelussa on, sitä enemmän aineistoa tarvitaan. Tyyppien määrää ei liioin välttämättä tiedetä edeltä käsin, joten otoskoon minimi on parhaiten määriteltävissä silloin, kun tyyppien määrä tiedetään ennalta.

Perkinsin ohje koskien otoskokoa liittyy myös keskusteluun kielen universaaleista ja niiden todentamisesta tilastollisin menetelmin. Typologin tutkimusmateriaali tulee pääasiassa nykyään puhutuista kielistä, koska historiallisista kielimuodoista on säilynyt vain vähän dokumentteja ja ne muodostavat erittäin suppean otoksen historian saatossa puhutuista kielistä. Tämä ei itsessään ole ongelma, mutta ongelmaksi voi tulla se, millaisia oletuksia käyttämäämme otantamenetelmään liittyy. Kuten yllä todettiin, typologiassa on tapana käyttää ositettua otantaa, jotta otoskielet olisivat toisistaan riippumattomia. Tällöin otokseen valitaan tyypillisesti vain yksi kieli kustakin kieliperheestä tai genuksesta. Mutta kun valitaan yksi kieli kustakin kieliperheestä, otoksen perusteella ei voi päätellä, miten lähisukukielet muuttuvat kieliperheiden sisällä, vaan tuloksena on ainoastaan synkroninen tilannekuva siitä, miten tutkittu kielellinen rakenne on jakautunut maailman kielissä tiettynä aikana. Jotta tutkija voi perustellusti tulkita, että tällainen synkroninen jakauma heijastaa kielellistä universaalia, hänen on oletettava, että rakennetyyppien jakaumat ovat saavuttaneet ns. tasapainojakauman (engl. *stationary distribution*). Tasapainojakauma tarkoittaa sitä, että muutokset yhdestä tyypistä toiseen tapahtuvat yhtä taajaan eri tyyppien kesken: koska kaikki muutokset

tyyppien välillä ovat yhtä yleisiä, synkroninen frekvenssijakauma ei enää muutu ajan myötä (Maslova 2000). Ongelmaksi muodostuu, ettei ole lainkaan selvää, miten tasapainojakauma voitaisiin todentaa.

Ratkaisuksi on tarjottu lähestymistapaa, jota kutsutaan *dynaamiseksi typologiaksi* (esim. Greenberg 1978; Croft 2003; Bickel 2015). Sillä on juurensa jo 1970-luvun typologisessa ajattelussa, mutta sitä kyettiin soveltamaan matemaattisesti ja tilastollisesti vasta 2000-luvulla (etenkin Maslova 2000). Dynaamisessa typologiassa luovutaan vaatimuksesta, että otoskielet olisivat toisistaan riippumattomia, eikä otantaa käytetä siihen, että suljettaisiin pois kielisukulaisuuden ja kielikontaktien sekoittava vaikutus. Tämä tarkoittaa sitä, että samasta kieliperheestä valitaan otokseen useita kieliä, jotta voidaan tutkia, millaisia diakronisia preferenssejä sukukielten kesken esiintyy. Kieliperheiden ja maantieteellisten alueiden vaikutusta voidaan sitten testata rakentamalla ne samaan tilastolliseen malliin testattavien lingvististen rakenteiden kanssa. Tällaiseen tilastolliseen tutkimukseen vaaditaan jo monimutkaisia monimuuttujamenetelmiä, kuten logistista regressiota tai yleistä monitasomallia (engl. *generalized multilevel modelling*). On syytä pitää mielessä, että dynaamisen lähestymistavan käyttö kasvattaa otoskokoa merkittävästi, koska kieliä valitaan sekä eri kieliperheistä että myös kieliperheiden sisältä. Tyypillinen otoskoko dynaamisessa typologiassa on noin 300–400 kieltä, usein enemmän (esim. Bickel 2015).

## 3.3. Kielimateriaalin hankinta

Kielimateriaalin hankintaan vaikuttavat oleellisesti tutkimuskysymys ja otoksen koko, ja tavallisesti materiaali hankitaankin samanaikaisesti otantaa tehdessä. Tyypillisin tutkimusmateriaali kielitypologiassa on yksittäisen kielen kielioppikuvaus tai tutkimuksen kohteena olevaa ilmiötä yksittäisessä kielessä käsittelevä tutkimus. Lisäksi sähköinen viestintä mahdollistaa vaivattoman yhteydenoton eri kielten asiantuntijoihin, jos omat lähteet eivät ole riittäviä.

Tutkimusmateriaalin saatavuudessa on kuitenkin suurta hajontaa. Suurista kielistä, kuten suomen kielestä, on olemassa useita

perusteellisia kuvauksia, kun taas pienistä ja uhanalaisista vähemmistökielistä ei välttämättä ole saatavilla muuta kuin yksittäisiä artikkeleita, jos niitäkään, ja ne riittävät vain harvoin lähteiksi omaan tutkimukseen. Eri kielikunnat ja maantieteelliset alueet poikkeavat toisistaan paljonkin siinä, miten syvällisesti ja laajasti niiden kieliä on tutkittu ja kuvattu. Etenkin Afrikassa ja Uudessa-Guineassa on paljon kieliä, joista ei ole vielä laadittu edes suppeaa kielioppikuvausta.

Tämä vinouma lähdemateriaalin saatavuudessa vaikuttaa siihen, mitä kieliä eri kielikunnista ja eri maantieteellisiltä alueilta voi ylipäätään valita otokseen. Tämän vuoksi lähdemateriaalin saatavuus on otettava huomioon otannassa alusta pitäen. Usein typologi joutuukin palaamaan otantaan ja vaihtamaan joitakin otoskieliä sen takia, että alun perin otokseen valittua kieltä ei kyetä analysoimaan, koska lähdemateriaali osoittautuu riittämättömäksi. Pahimmillaan lähteiden vaihteleva saatavuus voi johtaa siihen, että jotkin maantieteelliset alueet tulevat otoksessa aliedustetuiksi ja toiset taas vastaavasti yliedustetuiksi.

Kokeneella typologilla on paljon hiljaista tietoa siitä, mistä kielistä on ylipäätään saatavilla tutkimusmateriaalia, ja viime vuosina on kehitetty myös erittäin hyviä työkaluja helpottamaan materiaalin keräämistä. Erityisesti mainittakoon *Glottolog* (Hammarström, Forkel & Haspelmath 2019), joka on paitsi katalogi maailman kaikista kielistä, niiden sukulaissuhteista ja maantieteellisestä sijainnista myös viitetietokanta, joka sisältää noin 330 000 viitettä etenkin vähemmän tutkittujen kielten tutkimuskirjallisuuteen. Otantaa tehtäessä *Glottolog* on verraton työkalu, jonka avulla saa nopeasti käsityksen siitä, millaista tutkimuskirjallisuutta tietystä kielestä on saatavilla. Hyviä apuvälineitä otantaan ja tutkimusmateriaalin löytämiseksi ovat lisäksi *The Ethnologue* (Eberhard, Simons & Fennig 2019) ja WALS:n verkkoversio (Dryer & Haspelmath 2013). Varsinaista lähdemateriaalia on jo melko hyvin saatavilla avoimesti verkossa. Lisäksi Kansalliskirjaston finna.fi-palvelun kautta on saatavilla laadukkaita kuvauksia sadoista kielistä, mm. iso osa Mouton de Gruyterin julkaisemasta kirjasarjasta *Mouton Grammar Library*.

Typologit tekevät usein itsekin kenttätyötä, mutta pääasiassa typologista tutkimusta tehdään kuitenkin toisten kielentutkijoiden laatimien tutkimusten perusteella. Tämä ei kuitenkaan tarkoita sitä, että tutkija

sellaisenaan kopioisi lähdemateriaalissa olevan analyysin omaan tutkimukseensa. Typologi on toki paljolti riippuvainen lähdemateriaalin analyyseista, mutta kunnollinen lähdekritiikki on ensiarvoisen tärkeää, ja typologi muodostaa lähteiden pohjalta oman analyysinsa, mahdollisuuksien mukaan useita eri lähteitä samasta kielestä vertaillen. Tarkkuus analyysissa on tärkeää paitsi tutkimuksessa tehtävien yleistysten ja johtopäätösten kannalta myös sen tutkimuseettisenkin kysymyksen kannalta, että yksittäiset väärin analysoidut esimerkit voivat jäädä elämään kielitieteellisessä kirjallisuudessa vääriNä käsityksinä jostakin yksittäisestä kielestä (tutkimusetiikasta ks. Johdannon luku 4 tk.). Tutkijan on opittava arvioimaan lähteiden luotettavuutta, vertailtavuutta ja sitä, millaisesta teoreettisesta näkökulmasta kielen kuvaus on tehty. Lähteiden vertailtavuuteen liittyy esimerkiksi kysymys kirjoitetun ja puhutun kielen suhteesta. Kieliopit pohjautuvat tyypillisesti puhuttuihin muotoihin, mutta etenkin Euroopan kielten ja ylipäätään suurten kielten kielioppien taustalla vaikuttaa kirjoitettu yleiskieli. Tällaiset erot aineiston luonteessa voivat mahdollisesti vaikuttaa otoksesta tehtäviin typologisiin yleistyksiin (Himmelmann 2000, 10–11).

Lähteiden käytettävyyteen liittyy myös kysymys metakielestä, jolla ne on kirjoitettu. Kuten luvussa 2 todettiin, referenssikieliopit kirjoitetaan nykyisin yleisesti peruskielioppiteorian termein ja käsittein, jolloin ne ovat paljon laajemman tutkijajoukon käytettävissä kuin jonkin spesifin teorian formaalilla metakielellä kirjoitetut kieliopit. Tästä ääriesimerkkinä voisi mainita 1960–1970-luvuilla tagmemiikan piirissä tuotetut kieliopit, jotka ovat teoriaa tuntemattomalle tutkijalle erittäin vaikeaselkoisia.[3]

Kielioppikuvaukset eivät suinkaan ole ainut mahdollinen lähde kielten laajassa vertailussa. Esimerkiksi saman alkuperäistekstin käännöksiä eri kielille sisältävien rinnakkaiskorpusten käyttö on yleistynyt viime vuosina typologiassa. Tyypillisimpiä käytettyjä käännöksiä ovat raamatunkäännökset, elokuvatekstitykset ja EU-tekstit.[4] Rinnakkaistekstit avaavat uudenlaisia mahdollisuuksia vertailuun, ne mahdollistavat mm. tekstifrekvenssien vertailun, mutta niillä on omat rajoituksensa siinä, minkä tyyppisiin tutkimuskysymyksiin niitä voidaan käyttää. Lisäksi on muistettava, että rinnakkaistekstit ovat käännöstekstejä ja voivat tekstilajina poiketa ei-käännetystä kielenkäytöstä (ks. Käännöstiede

ja sen menetelmät tk.). Kun rinnakkaiskorpuksen avulla vertailtavien kielten määrä kasvaa hyvin suureksi, jopa useisiin satoihin ja yli tuhanteenkin kieleen (engl. *massively parallel corpora*), tarvitaan erityisiä metodeja, joita on kehitetty näin laajaan korpuspohjaiseen vertailuun (esim. Cysouw & Wälchli 2008).

Myös kyselylomakkeita käytetään typologiassa jonkin verran. Ne voivat olla hyödyllinen tapa kielimateriaalin hankintaan, etenkin jos tutkittavasta ilmiöstä on vaikea saada tietoa kielioppikuvausten tai tekstien avulla (ks. esim. Dahl 1985; Siewierska 1998; myös Kysely tk.) Kyselylomakkeet mahdollistavat systemaattisen datan keräämisen ja soveltuvat hyvin laadittuina erilaisiin tutkimuskysymyksiin joko ensisijaiseksi tai täydentäväksi datankeruumenetelmäksi. Kyselylomakkeissa on kuitenkin omat metodologiset ongelmansa: niillä kerätty data ei ole luonnollista kielenkäyttöä, varsinkaan jos kysely koostuu metakielestä käännettävistä lauseista; kuvista tai videoista koostuvissa kyselyissä ongelma on pienempi. Tällaisten elisitaatioon yleisesti liittyvien ongelmien lisäksi usein vielä lomakkeen täyttävä tutkija on ylimääräisenä välikätenä kyselylomakkeen laatineen tutkijan ja kielikonsultin välillä (ks. esim. Croft 2003, 29). Kyselylomakkeita, kuten mitä tahansa muutakin datankeruumenetelmää, käytettäessä on pidettävä kirkkaana mielessä, mihin niillä kerätty data soveltuu ja mihin ei.

Typologisen tutkimuksen lähteitä ja niiden asemaa kieliotoksessa auttaa ymmärtämään **dokulektin** käsite. Dokulektillä viitataan jonkin konkreettisen dokumentin esittämään kielimuotoon. Tällaisia dokumentteja ovat esimerkiksi referenssikieliopit, sanakirjat ja korpukset. Esimerkiksi *Iso suomen kielioppi* (Hakulinen, Vilkuna, Korhonen, Koivisto, Heinonen & Alho 2004) edustaa yhtä dokulektiä ja *Finnish: A Comprehensive Grammar* (Karlsson 2018) toista. Typologin kieliotos koostuu siis itse asiassa dokulekteistä, ja kieltenväliset yleistykset tehdään dokulektien pohjalta. Typologin tutkimuskysymys ei koske ensisijaisesti tutkittavaa ilmiötä yksittäisessä kielessä vaan otokseen valikoituneiden dokulektien pohjalta piirtyvää kieltenvälistä yleistystä. Tällä seikalla voidaan ajatella olevan myös tutkimuseettisiä ulottuvuuksia: typologin tehtävänä ei ole reifioida (eli tehdä olevaksi) tai lokeroida kieliä, mitä on hyvä korostaa väärinymmärrysten välttämiseksi.

Kielitypologiassa ei tyypillisesti olla suoraan tekemisissä kielenpuhujien kanssa, joten puhujien suostumukseen ja anonymiteettiin liittyvät tutkimuseettiset kysymykset koskevat typologiaa yleensä vain epäsuorasti. Toki toisten tutkijoiden keräämää aineistoa käytettäessä on kiinnitettävä huomiota aineiston käyttöehtoihin ja toimittava niiden mukaisesti. Kyselykaavakkeilla aineistoa kerättäessä nämä kysymykset ovat suoremmin osa tutkimusprosessia.

## 3.4. Analyysi ja luokittelu

Kun otos on laadittu ja otoskielistä on hankittu lähdemateriaali, seuraava askel on kielimateriaalin analysointi ja tarkasteltavana olevassa kielenpiirteessä havaittavan kieltenvälisen variaation luokittelu eri (rakenne)tyyppeihin. *Kielenpiirteellä* tarkoitetaan tässä siis tarkasteltavana olevaa, tyypillisesti kielen rakenteeseen liittyvää ilmiötä, kuten sanan morfologista perusrakennetta, transitiivisen lauseen hallitsevaa sanajärjestystä tai vaikkapa negaation ilmaisua. Typologit pyrkivät sekä kuvaamaan että selittämään kielenpiirteissä ilmenevää variaatiota, ja luokittelu on keskeinen vaihe tässä prosessissa. Esimerkiksi morfologisen perusrakenteen eri tyyppejä ovat luvussa 2 mainitut agglutinoivat, fusionaaliset ja isoloivat rakenteet, kun taas transitiivisen lauseen hallitsevan sanajärjestyksen tavallisia tyyppejä ovat ne eri järjestykset, joihin lauseen osallistujat ('subjekti' ja 'objekti') sekä verbi voidaan asettaa, kuten SVO, SOV tai VOS. Analyysi ja luokittelu kulkevat rinta rinnan, ja luokittelu tarkentuu sitä mukaa, kun uusia otoskieliä analysoidaan. Tällöin tyyppien määrittely voi päivittyä ja myös uusia tyyppejä voidaan tarvittaessa lisätä.

Luokittelusta ei koskaan tule aukoton, sillä jokin uusi otoskieli voi poiketa muista niin paljon, ettei sitä voida luokitella samaan rakennetyyppiin muiden kielten kanssa. Tyypit ovat kieltenvälisiä yleistyksiä, joten jonkin ilmiön löytyminen yhdestä tai kahdesta kielestä ei vielä oikeuta omaa tyyppiään. Toisaalta voi olla, että jossakin menneisyydessä puhutussa tai tulevaisuudessa puhuttavassa kielessä on esiintynyt tai esiintyy rakenne, jota ei esiinny missään otoskielessä tai edes nykyään

puhuttavissa kielissä. Esimerkiksi 1970-luvulla kielitieteessä uskottiin, että lauseen pääasiallisessa sanajärjestyksessä objekti ei koskaan esiintyisi ennen subjektia ja verbiä (ns. OSV- ja OVS-sanajärjestykset). Tilanne kuitenkin muuttui, kun kenttätutkijat alkoivat julkaista enemmän kuvauksia Amazonin ja Uuden-Guinean kielistä. Etenkin Derbyshiren (1979; 1985) kuvaukset hixkaryanasta vakuuttivat epäilijät. Hixkaryana on yksi harvoista maailman kielistä, joiden perussanajärjestys on OVS (ks. esim. 2). Huomattakoon, että kun rakennetyyppi on yksi looginen mahdollisuus rajattujen mahdollisuuksien joukossa, kuten OVS sanajärjestysten joukossa, voidaan puhua tyypistä, vaikka se esiintyisi vain yhdessä kielessä.

2)

| | | |
|---|---|---|
| *Toto* | *y-ahosɨ-ye* | *kamara.* |
| mies | 3A.3P-tarttua-PST.CMPL | jaguaari |

'Jaguaari nappasi miehen.' (Derbyshire 1985, 32)

Tämän esimerkin yhteydessä voimme myös todeta, että typologisessa tutkimuksessa käytetään nykyisin yleisesti ns. Leipzigin glossaussääntöjä (Leipzig Glossing Rules),[5] joiden avulla voidaan tehdä erilaisten, lukijalle tuntemattomien kielten rakenteet näkyviksi ja ymmärrettäviksi. Esimerkkien valintaan liittyy myös tutkimuseettisiä kysymyksiä: on hyvä välttää mahdollisuuksien mukaan esimerkkejä, joiden sisältö saatetaan kokea loukkaavana. Lisäksi esimerkkien valinnalla voidaan myös vaikuttaa siihen, miten eksotisoivaa kuvaa eri kieliyhteisöistä pidetään yllä. Esimerkiksi Daniel Everettin väitteet koskien pirahã-kieltä ja sen puheyhteisön kulttuuria ovat saaneet osakseen kritiikkiä tästä näkökulmasta (ks. Everett 2004).

Analyysivaiheessa huomio kiinnittyy etenkin kieltenvälisiin yhtäläisyyksiin ja eroihin. Näiden perusteella kielimateriaalia jäsennellään ja aineistossa esiintyvistä rakenteista tehdään typologinen luokittelu. Analyysiprosessin yksityiskohdat riippuvat pitkälti siitä, miten syvällisesti kukin otoskieli analysoidaan. Mikäli kielellinen ilmiö on melko

suoraviivaista analysoida, riittää, että kirjaa itselleen luokittelun määritelmät ja periaatteet, joita kielten analyysissa sovelletaan ja joita sen kuluessa mahdollisesti tarkennetaan. Tällöin voi otoskieliä analysoitaessa samanaikaisesti rakentaa tietokantaa (tietokannoista ks. Laadullinen aineistopohjainen kielentutkimus tk.), johon kunkin otoskielen ominaisuudet koodataan, eikä tekstimuotoon auki kirjoitettua analyysia tarvitse välttämättä tuottaa kuin niistä tapauksista, joiden analyysi vaatii eri vaihtoehtojen punnitsemista. Esimerkiksi Sinnemäen (2014a) noin 700 kielen analyysi objektin differentiaalisesta sijanmerkinnästä noudatteli pääosin tätä analyysitapaa: objektin sijanmerkinnän tyypit analysoitiin ja kirjattiin tietokantaan lyhyin sanallisin kuvauksin, minkä pohjalta varsinainen luokittelu sitten koodattiin tietokantaan ja perusteellisempi, auki kirjoitettu analyysi tehtiin vain niistä kielistä, joiden luokittelu oli vaativampaa. Toisaalta, jos luokiteltava kielellinen ilmiö vaatii syvällisempää analyysia, analyysin aukikirjoittaminen voi olla välttämätöntä jokaisen otoskielen kohdalla. Esimerkiksi Sinnemäen (2014b) analyysi sijanmerkinnän ja sanajärjestyksen kompleksisuudesta vaati perusteellisen 2–3 sivun tekstimuotoisen analyysin kustakin 50 otoskielestä.

Tunnetuimpia typologisia luokitteluja ovat sanajärjestystypologia ja jo luvussa 2 mainittu klassinen morfologinen typologia. Siitä, millaisia luokitteluja typologiassa on esitetty näiden lisäksi, saa hyvän yleiskuvan WALS-tietokantaa selaamalla. Jokainen WALS:n karttaluku sisältää typologisen luokittelun, jonka eri tyyppien maantieteellisiä jakaumia sitten esitellään kartalla. On hyvä todeta, että kielten rakenteiden ja ominaisuuksien luokittelu ei ole koskaan itse tarkoitus vailla muuta päämäärää, vaan se on yksi välttämätön vaihe tutkimusprosessissa. Luokittelun on oltava teoreettisesti perusteltu ja motivoitu. Typologinen tutkimus pyrkii paitsi kuvaamaan myös ymmärtämään kieltenvälistä variaatiota. Luokittelussa tulee siis pyrkiä kiinnittämään huomiota seikkoihin, jotka ovat relevantteja kieltenvälisten yhtäläisyyksien ja eroavaisuuksien selittämisessä. Esimerkiksi negaation ilmaisu voitaisiin jakaa tyyppeihin sen mukaan, miten monta vokaalia negaatiota ilmaisevassa morfeemissa esiintyy. Tällainen luokittelu ei kuitenkaan auttaisi ymmärtämään negaatiota eikä tarjoaisi vastauksia mihinkään teoreettisesti kiinnostavaan negaatiota koskevaan tutkimuskysymykseen. Sen sijaan esimerkiksi

luokittelu tyyppeihin sen mukaan, miten kieltolauseet eroavat rakenteeltaan vastaavista myöntölauseista, voidaan yhdistää negaation funktionaalisiin ominaisuuksiin, minkä avulla voidaan löytää motivaatioita eri rakennetyypeille (ks. Miestamo 2005).

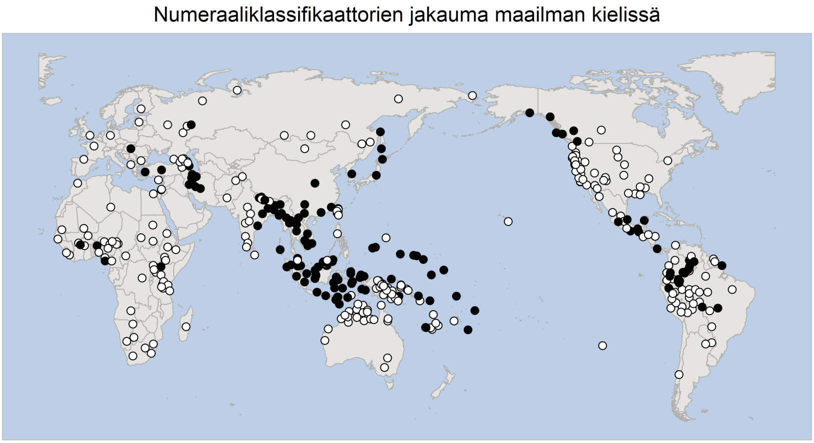
## 3.5. Rakennetyyppien jakaumien tarkastelu

Luokittelu on ensimmäinen vaihe, kun kieltenvälisestä variaatiosta tehdään yleistyksiä. Luokittelun pohjalta voidaan sitten siirtyä tarkastelemaan tutkittavan kielenpiirteen eri rakennetyyppien yleisyyttä, niiden genealogista (eli kielikuntakohtaista) ja maantieteellistä jakautumista sekä mahdollisia korrelaatioita muiden kielenpiirteiden kanssa. Yleensä tyyppien voidaan havaita esiintyvän epätasaisesti tarkasteltaessa maailmanlaajuisia frekvenssejä tai niiden maantieteellisiä jakaumia, ja typologin tehtävänä on selvittää, mistä tämä johtuu.

Kielenpiirteiden epätasainen jakauma voi johtua genealogisista tai maantieteellisistä syistä eli siitä, että jokin tyyppi on yleisempi kuin muut tietyssä kieliperheessä tai tietyllä maantieteellisellä alueella. Näissä tapauksissa voidaan sanoa, että frekvenssien epätasainen jakauma johtuu historiallisista tapahtumista yksittäisissä kieliperheissä, joissa tyypin yleisyys johtuu sen periytymisestä, tai yksittäisillä alueilla, joissa sama tyyppi on levinnyt kielestä toiseen kielikontaktin vaikutuksesta. Tutkimus on osoittanut, että suhteellisen harvan kielenpiirteen kohdalla eri tyypit ovat jakautuneet tasaisesti eri maanosissa, mikä kertoo kielikontaktien ja osittain myös kielisukulaisuuden vaikutuksesta tyyppien levinneisyydessä.

Kielenpiirteiden levinneisyyttä tarkastellaan typologiassa yleensä myös käyttämällä karttoja, joiden avulla tyyppien maantieteellisiä jakaumia on mahdollista helposti hahmottaa. Yllä mainittu WALS-kartasto on hyvä esimerkki toisaalta siitä, miten typologit ovat alkaneet kiinnittää yhä enemmän huomiota maantieteellisiin jakaumiin, ja toisaalta siitä, miten tärkeä työkalu kartat ovat typologille. Siinä missä maailman poliittiset kartat piirretään yleensä Atlantti-keskisesti, typologiassa kartat piirretään usein Tyynimeri-keskisesti. Tämän taustalla on ajatus siitä, että

joidenkin piirteiden maailmanlaajuiset jakaumat voivat heijastaa muinaisia väestön liikkeitä ja jopa ihmislajin leviämistä Afrikasta muihin maanosiin. On esimerkiksi todettu, että Tyynenmeren ympärillä puhuttavissa kielissä on samankaltaisuutta useissa kielenpiirteissä, joiden kohdalla muualla maailmassa esiintyy enemmän variaatiota (Nichols 1992); tästä esimerkkinä on numeraaliklassifikaattorien jakauma kuviossa 1.




Kuvio 1. Kartta numeraaliklassifikaattorien (Gil 2013) esiintymisestä maailman kielissä (valkoinen = ei esiinny; musta = esiintyy).[6]

Kielisukulaisuuden ja kielikontaktien lisäksi – ja myös niistä riippumatta – on mahdollista, että rakennetyyppien maailmanlaajuiset frekvenssijakaumat viittaavat kielelliseen universaaliin. Universaalit ovat tavallisimmin tilastollisia preferenssejä, siis havaintoja siitä, että tietyn kielenpiirteen jokin rakennetyyppi on muita yleisempi. Esimerkiksi lauseen sanajärjestyksen suhteen subjekti yleensä edeltää objektia, mikä tarkoittaa sitä, että rakennetyyppi ”subjekti edeltää objektia” on yleisempi kuin tyyppi ”objekti edeltää subjektia” (esim. Bickel, Witzlack-Makarevich, Choudhary, Schlesewsky & Bornkessel-Schlesewsky 2015). Tämä on esimerkki tilastollisesta universaalista, mutta universaalit voivat olla myös absoluuttisia. Absoluuttiset universaalit koskevat kaikkia ihmiskieliä (esim. kaikilla kielillä on CV-tavuja). Teoreettisesti

kiinnostavia absoluuttisia universaaleja ei tutkimuksessa ole juurikaan löydetty, ja nekin, joiden on väitetty olevan absoluuttisia, ovat saaneet osakseen kovaa kritiikkiä (mm. Evans & Levinson 2009). Absoluuttisten universaalien löytäminen typologisin työkaluin voi olla jopa mahdotonta (Piantadosi & Gibson 2014). Funktionaalis-typologisessa tutkimuksessa tehdään tyypillisesti tilastollisia yleistyksiä kielistä ja oletetaan lingvististen rakenteiden jakautuvan probabilistisesti eli esiintyvän eri todennäköisyyksillä (ks. Bickel 2007).

Typologisiin yleistyksiin liittyy myös yleinen harhakäsitys koskien vastaesimerkkejä. Toisinaan etenkin ei-typologit voivat ajatella niiden voivan kumota typologiset yleistykset. Yksittäiset vastaesimerkit ovat potentiaalisesti ratkaiseva ongelma kuitenkin vain absoluuttisille universaaleille, sillä vastaesimerkit otetaan huomioon tilastollisessa testaamisessa. Vasta siinä tapauksessa, että uusia vastaesimerkkejä löytyy merkittävä määrä eri kielikunnista ja eri alueilta otoksen ulkopuolelta, voidaan tilastollinen typologinen yleistys kumota – aivan samoin periaattein kuin muissakin tieteissä.

## 3.6. Typologiset korrelaatiot

Korrelaatiot ovat kieltenvälisiä yleistyksiä kahden tai joskus useamman loogisesti toisistaan riippumattoman kielenpiirteen välisestä suhteesta. Riippumattomuus tarkoittaa tässä sitä, että kielenpiirteiden määrittelyt eivät ole keskenään päällekkäisiä vaan määritelmien avulla rajataan tutkittavaksi eri kielenpiirteet, kuten vaikkapa sanojen keskinäinen järjestys ja sanojen taivutus. Myös korrelaatiot voivat päteä kaikkiin kieliin (= absoluuttiset universaalit) tai ne voivat päteä suureen osaan kieliä (= tilastolliset universaalit). Typologiset korrelaatiot on perinteisesti esitetty implikationaalisina universaaleina eli väitteinä, jotka kertovat kielenpiirteiden välisestä riippuvuussuhteesta implikaation muodossa $p > q$ (luetaan ”jos p, niin q”) (Greenberg 1963; ks. myös Miestamo 2010). Tunnetuimpia tällaisia implikationaalisia universaaleja ovat sanajärjestystä koskevat yleistykset, joissa eri lause- ja lauseketyyppien sanajärjestysominaisuudet ovat riippuvuussuhteessa keskenään. On

esimerkiksi havaittu, että jos kielen hallitseva sanajärjestys on VSO, sen adpositiot ovat yleensä prepositioita, ja jos taas kielen hallitseva sanajärjestys on SOV, sen adpositiot ovat postpositioita (ks. Greenberg 1963; Dryer 1992). Tässä on itse asiassa kyse kahdesta erillisestä implikationaalisesta universaalista:

3) VSO > prepositioita
SOV > postpositioita

Eri rakenteiden sanajärjestysominaisuuksien välillä on havaittu laajempia korrelaatioita niin, että voidaan puhua toisaalta VO- ja toisaalta OV-kielistä, joissa verbin ja objektin järjestyksen kanssa korreloivat mm. taulukossa 1 lueteltujen konstituenttien keskinäiset järjestykset.

| verbi, objekti | VO | OV |
|---|---|---|
| adpositio, substantiivi | AdpN (=prepositiot) | NAdp (=postpositiot) |
| omistettu (=N), omistaja (=G) | NG | GN |
| substantiivi, relatiivilause | NRel | RelN |

Taulukko 1. Sanajärjestyskorrelaatioita.

Kun implikationaalisissa universaaleissa yhdistyy vähintään kaksi toisistaan loogisesti riippumatonta kielenpiirrettä, ne ovat askel partiaalisesta, vain yhtä kielen ilmiötä koskevasta typologiasta kohti holistisempaa typologiaa. Voidaan ajatella, että partiaalisen ja holistisen typologian välillä on jatkumo: voimme yhdistää korrelaatioiden kautta kielenpiirteitä toisiinsa ja päästä lähemmäs kielen rakenteen kokonaiskuvaa. Kielitypologian yksi keskeinen tavoite onkin selvittää, millä tavoin ja missä määrin kielenpiirteet korreloivat keskenään. Kuitenkin sellainen typologinen parametri tai parametrien joukko, joka jotenkin kokonaisvaltaisesti luonnehtisi kutakin yksittäistä kieltä, näyttää nykytietämyksen valossa utopistiselta tavoitteelta. Tällaisena voidaan pitää esimerkiksi 1800-luvulla esiintynyttä ajatusta, että kielen morfologinen tyyppi kertoisi kielestä jotakin yleisempää.

Implikationaaliset hierarkiat ovat useampien kielenpiirteiden välisiä riippuvuusketjuja, siis oikeastaan peräkkäin aseteltuja implikationaalisia universaaleja. Esimerkki implikationaalisesta hierarkiasta on lukuhierarkia, jonka esitämme tässä yksinkertaistetussa muodossa:

4) yksikkö < monikko < duaali < triaali

Jos kielessä esiintyy jokin tällaisessa hierarkiassa mainittu ominaisuus, hierarkia kertoo, että silloin myös kaikki tästä vasemmalle olevat ominaisuudet esiintyvät. Ominaisuuden esiintyminen siis implikoi kaikki siitä nuolen suuntaan olevat ominaisuudet. Lukuhierarkian tapauksessa, jos esimerkiksi todetaan, että kielessä on (kahta tarkoitetta ilmaiseva) duaalikategoria, voidaan hierarkian mukaisesti päätellä, että siinä on myös monikko ja yksikkö. Hierarkian mukaan duaalin läsnäolo kielessä ei kuitenkaan kerro mitään siitä, onko kielessä mahdollisesti (kolmea tarkoitetta ilmaiseva) triaali.

Erityisen keskeiseksi kielissä vaikuttavaksi hierarkiaksi on osoittautunut elollisuushierarkia. On esitetty, että se säätelee erilaisia kieliopin ilmiöitä eri kielissä ja kiteyttää esimerkiksi objektin merkintää koskevan yleistyksen (Croft 2003; Sinnemäki 2014a). Tämän yleistyksen mukaan jos hierarkiassa alempana esiintyvä referentti merkitään objektisijalla jossakin kielessä, myös kaikki hierarkiassa ylempänä esiintyvät referentit merkitään objektisijalla:

5) ihminen < muu elollinen < eloton

Hierarkia kertoo myös sen, että elolliset objektit merkitään sijalla useammin kuin elottomat objektit.

Vaikka implikationaaliset universaalit ovatkin olleet suosittu menetelmä typologisten korrelaatioiden esittämiseen, niiden käytössä on tiettyjä ongelmia (Cysouw 2005). Yksi ongelma liittyy siihen, että jotkin implikationaalisten universaalien kannalta kiinnostavat jakaumat eivät tilastotieteen näkökulmasta ole lainkaan kiinnostavia. Oletetaan esimerkiksi kaksi kuvitteellista sadan kielen jakaumaa (ks. taulukko 2), jotka esittävät piirteiden A ja B esiintymistä. Perinteisessä typologiassa

molemmat jakaumat ovat periaatteessa yhtä kiinnostavia, koska ne viittaavat piirteiden A ja B väliseen korrelaatioon: kummassakaan otoksessa ei esiinny lainkaan rakennetyyppien A+ ja B– yhdistelmää.

| (a) | A+ | A– | (b) | A+ | A– |
|---|---|---|---|---|---|
| B+ | 15 | 70 | B+ | 25 | 50 |
| B– | 0 | 15 | B– | 0 | 25 |
| $p = 0.12$ | | | $p = 0.0003$ | | |

Taulukko 2. Implikationaaliset universaalit ja tilastollinen testaaminen. (+ = piirre esiintyy kielessä; – = piirre ei esiinny kielessä; *p* = Fisherin tarkalla testillä saatu todennäköisyys jakaumasta.)

Nämä jakaumat voitaisiin esittää kahden implikationaalisen universaalin avulla: A+ > B+ (eli jos kielessä esiintyy piirre A, siinä esiintyy myös piirre B) ja B– > A– (eli jos kielessä ei esiinny piirrettä B, siinä ei esiinny myöskään piirrettä A). Kuitenkin jos näitä jakaumia testataan tilastollisesti (tässä Fisherin tarkalla testillä), ainoastaan jakauma (b) poikkeaa odotetusta jakaumasta tilastollisesti merkitsevällä tavalla ($p < 0.05$, jota käytetään yleisesti kielentutkimuksessa tilastollisen merkittävyyden rajana), kun taas jakauma (a) ei ($p > 0.05$). Näin ollen vain jakauma (b) on tilastollisessa mielessä kiinnostava. Vaikka perinteisessä typologiassa on usein käytetty tilastollisia menetelmiä, implikationaalisia universaaleja ei ole aina todennettu tilastollisesti vaan ainoastaan frekvenssien perusteella. Tämä lienee yksi syy siihen, miksi modernissa typologiassa typologisia korrelaatioita ei esitetä enää kovin usein (pelkkien) implikationaalisten universaalien vaan edistyneempien tilastollisten menetelmien avulla. Erityisesti monimuuttujamenetelmien, kuten erilaisten regressiomenetelmien, käyttö onkin lisääntynyt huomattavasti typologiassa (mm. Cysouw 2010; Sinnemäki 2011; Bickel 2015). Näiden menetelmien etuna on se, että kielisukulaisuuden ja kielikontaktien vaikutuksen suuruutta voidaan erikseen testata tilastollisesti. Esimerkiksi Sinnemäki (2011) osoitti, että kielenpiirteen jakaumaan voivat vaikuttaa niin muut kielenpiirteet kuin myös kielisukulaisuus ja kielikontaktit, mutta kuitenkin toisistaan riippumatta.

Viime vuosina typologit ovat kiinnostuneet myös kielenpiirteiden suhteesta kielenulkoisiin tekijöihin. Kielenpiirteiden jakaumia on tutkittu esimerkiksi suhteessa puheyhteisön kokoon (Lupyan & Dale 2010; Sinnemäki & Di Garbo 2018), kielen puhuma-alueen ilmastoon (Everett, Blasi & Roberts 2015) ja geneettisiin tekijöihin (Dediu & Ladd 2007). Kielentutkimuksessa on perinteisesti ajateltu, että kielenpiirteet eivät vaihtele tällaisten tekijöiden suhteen, ja käytännössä 1900-luvun loppuun asti tutkimuksessa vältettiin tällaisia tutkimusaiheita muutamia yksittäisiä poikkeuksia lukuun ottamatta. 2000-luvun alusta lähtien myös kielenulkoiset sosiaaliset, ekologiset ja biologiset tekijät ovat herättäneet kiinnostusta ja keskustelua typologien kesken. Kun suuntaamme katseemme kielenpiirteiden ja kielenulkoisten tekijöiden suhteeseen, olemme tekemisissä kieltenvälisten yleistysten selittämisen kanssa. Seuraavassa luomme lyhyen katsauksen selittämiseen kielitypologiassa.

## 3.7. Tulosten selittäminen

Tutkimusprosessin viimeinen vaihe on tulosten selittäminen. Kuten yllä todettiin, typologista tutkimusta tehdään tavallisimmin funktionalistisessa viitekehyksessä. Selityksiä etsitään siis tyypillisesti kielenulkoisista tekijöistä, jotka vaikuttavat ilmiön syntyyn tai sen mukautumiseen ja sitä kautta kielten synkroniseen tilaan. Perinteiset funktionaaliset selitykset ammentavat semantiikasta, pragmatiikasta ja yleisistä kognitiivisista tekijöistä, kuten prosessointiin ja havaintoon liittyvistä seikoista. Viime aikoina on kiinnitetty myös enenevässä määrin huomiota kielen ekologiaan eli niihin lingvistisiin, sosiaalisiin, kulttuurisiin, historiallisiin, maantieteellisiin ja fyysisiin ympäristöihin, joissa kutakin kieltä käytetään ja joiden kanssa kieli on vuorovaikutuksessa (esim. Haugen 1972; Mühlhäusler 1996).

Yleisimpiä funktionaalisia selitysperusteita ovat ekonomia ja ikonisuus (ks. erityisesti Croft 2003). Ekonomia tarkoittaa kielellisten rakenteiden ja prosessoinnin taloudellisuutta: suositaan sellaisia rakenteita, jotka päästävät puhujan mahdollisimman vähällä vaivalla. Ekonomia on ollut suosittu selitysperuste kielentutkimuksessa jo pitkään (esim.

Jespersen 1922, 261–264; Zipf 1949; Martinet 1955). Ikonisuus taas pyrkii mahdollisimman hyvään vastaavuuteen muodon ja merkityksen välillä ja toimii näin ekonomian kanssa kilpailevana motivaationa ja sen vastavoimana (esim. Haiman 1980). Esimerkki ekonomian ja ikonisuuden välisestä kilpailusta on morfologisen typologian eri tyyppien välinen kamppailu: ikonisuus pyrkii luomaan ja säilyttämään agglutinatiivista rakennetta, jossa jokaista merkitystä vastaa yksi muoto, kun taas ekonomisuus pyrkii rikkomaan tätä ideaalia mm. lyhentämällä muotoja ja fuusioimalla ja assimiloimalla niitä toisiinsa. Ekonomia vaikuttaa etenkin kielen yleisimpien sanojen ja muotojen ominaisuuksiin, ja muotojen frekvenssiin nojaavat ekonomiapohjaiset selitykset ovat hyvä esimerkki käyttöpohjaisen lähestymistavan keskeisyydestä typologiassa (ks. Greenberg 1966; Haspelmath 2005; 2008).

Myös kielen prosessoinnilla tai omaksumisella on pyritty selittämään useita typologisia yleistyksiä (ks. Sinnemäki 2014c). Etenkin edellisessä luvussa mainittuja sanajärjestyskorrelaatioita on selitetty kielen prosessoinnilla. Hyvin yleisellä tasolla nämä selitykset perustuvat siihen, että kun samanlaisessa syntaktisessa asemassa olevat elementit sijoittuvat johdonmukaisesti samaan järjestykseen kussakin kielessä, kielen syntaktinen säännönmukaisuus lisääntyy ja kielen prosessointi helpottuu. Tunnetun tulkinnan näistä selitysperusteista on esittänyt Dryer (1992), ja vaihtoehtoisen prosessointimallin tarjoaa Hawkins (2004).

Poikkeavat rakenteet selitetään tyypillisesti historiallisilla tekijöillä. Niiden esiintyminen voidaan selittää joko kielikontakteilla tai poikkeavalla innovaatiolla satunnaisissa kielikunnissa. Esimerkiksi SVO-kielet, joissa on postpositioita, esiintyvät tyypillisesti sellaisilla maantieteellisillä alueilla, joilla eri sanajärjestystyyppejä edustavat kielet kohtaavat. Suomi on tästä hyvä esimerkki: suomalais-ugrilaisilla kielillä transitiivisen päälauseen hallitseva sanajärjestys on tavallisimmin SOV ja adpositiot ovat postpositioita. Suomessa transitiivisen päälauseen hallitseva sanajärjestys on kuitenkin muuttunut SVO:ksi, koska suomi on ollut pitkään kontaktissa germaanisiin, balttilaisiin ja slaavilaisiin kieliin, joilla on SVO-sanajärjestys ja prepositioita.

## 4. Yhteenveto: typologia kielitieteen kentässä

Kielitypologia on maailman kielten vertailevaa tutkimusta, mutta se ei toki ole ainut kielentutkimuksen ala, jossa tehdään kieltenvälistä vertailua tai jossa pyritään sanomaan jotakin yleistä ihmiskielestä. Yksi typologian lähialoista on luonnollisesti kontrastiivinen kielentutkimus, jossa myös verrataan kieliä toisiinsa ja pohditaan muun muassa, millä edellytyksillä kielten vertailu on mahdollista. Nämä alat kuitenkin eroavat toisistaan sekä tavoitteiltaan että menetelmiltään. Siinä missä kielitypologiassa tutkitaan systemaattisesti kieltenvälisiä eroja ja samankaltaisuuksia hyvinkin erilaisissa kielissä, kontrastiivisessa kielentutkimuksessa keskitytään tyypillisesti vertailemaan kahta kieltä, jotka ovat joko toisilleen sukua tai joita puhutaan maantieteellisesti samalla alueella. Kielitypologian perimmäinen tarkoitus on ymmärtää paremmin ihmiskielen luonnetta, joten sen tavoitteet ovat teoreettiset ja liittyvät pitkälti perustutkimukseen. Kontrastiivisen kielentutkimuksen tavoitteet ovat selvemmin soveltavia, ja tällainen tutkimus palvelee usein esimerkiksi kielenopetuksen tarpeita.

Kielitypologian lähialaksi voidaan katsoa myös (formalistinen) kielioppiteoria, sillä se on typologian tapaan kiinnostunut kielen universaaleista. Toisinaan onkin ajateltu, että kielitypologia ja chomskylainen universaalikielioppi muodostavat saman kolikon eri kääntöpuolet. Näiden alojen lähtökohdat ovat kuitenkin hyvin erilaiset, vaikka niiden kiinnostuksen kohteet osittain risteävätkin. Tässä artikkelissa olemme tarkastelleet etenkin funktionaalis-typologista lähestymistapaa kielentutkimukseen, ja tämän alan teoreettiset lähtökohdat ovat kielen käytössä ja kielen funktiossa maailman eri kielissä. Etenkin generatiivisessa kielitieteessä lähtökohta on kuitenkin kielen oletetussa biologisessa perustassa ja universaalissa kielikyvyssä, ja sen parissa on ollut tapana tutkia yhtä kieltä (usein englantia) syvällisesti toisin kuin kielitypologiassa, jossa on tapana tutkia kymmeniä tai jopa satoja kieliä samanaikaisesti. Bickel (2007) on jopa esittänyt, että funktionaalis-typologinen kielentutkimus on osa ihmiskulttuurin tutkimusta (laajasti ymmärrettynä), kun taas universaalikielioppi on osa kognitiotiedettä. Todettakoon

vielä, että kielen universaaleja ominaisuuksia tutkitaan myös monien kognitiivis-funktionalistisesti suuntautuneiden kielioppiteorioiden parissa (esim. langackerilainen kognitiivinen kielioppi ja goldbergiläinen konstruktiokielioppi; ks. Kieliopin tutkimus ja kielioppiteoriat tk.), ja nämä lähestymistavat ovat teoreettisesti hyvinkin yhteensopivia typologis-funktionalistisen ajattelun kanssa.

Kielitypologialle läheisimmät tutkimusalat löytyvät kuitenkin ns. diversiteettilingvistiikan piiristä. Tällä termillä on viime aikoina alettu kutsua toisiinsa läheisessä suhteessa olevia kielitieteen aloja, joiden yhteisenä tavoitteena on maailman kielellisen diversiteetin kartoittaminen ja ymmärtäminen. Näitä aloja ovat kielitypologian lisäksi vähemmän tutkittujen kielten dokumentointi ja kuvaus, antropologinen lingvistiikka, historiallis-vertaileva kielentutkimus, areaalilingvistiikka ja kielikontaktien tutkimus (ks. ko. lukuja tk.). Erityisesti kielitypologia ja deskriptiivinen kielentutkimus ovat läheisessä suhteessa. Kuten jo johdannossamme totesimme, niiden välillä vallitsee riippuvuussuhde: siinä missä deskriptiivinen kielentutkimus tuottaa tietoa typologeille aiemmin heikosti tunnetuista kielistä ja niiden rakenteista, kielitypologinen tieto auttaa kenttälingvistiä suhteuttamaan käsillä olevan kielen rakenteen analyysin siihen, mitä maailman kielistä tiedetään, ja näin ymmärtämään tutkimansa kielen ilmiöitä paremmin. Näidenkin alojen lähtökohdat ja menetelmät eroavat kuitenkin toisistaan. Deskriptiivinen kielentutkimus pyrkii ymmärtämään ja kuvaamaan yksittäisen kielen kielioppia, ja tyypillisesti kunkin kielet kategoriat on kuvattava niiden omilla ehdoilla (ks. luku 2). Kielitypologiassa typologi pyrkii luomaan käsitteitä, joiden avulla kieltenvälinen vertailu olisi mahdollista huolimatta yksittäisten kielten kategorioiden erilaisuudesta. Kielitypologi ei siis ensisijaisesti pyri sanomaan mitään jonkin yksittäisen kielen ominaisuuksista, vaan hänen tutkimuskysymyksensä koskevat kieltenvälistä vaihtelua – kielten erilaisuutta ja samuutta ja niiden taustalla vaikuttavia tekijöitä.

Tässä artikkelissa olemme esitelleet kielitypologiaa yhtenä lähestymistapana kielentutkimukseen. Esittelimme kielitypologian keskeisiä teoreettisia taustoja, sen tutkimushistorian pääpiirteet sekä typologisen tutkimusprosessin eri vaiheet. Lisäksi olemme käsitelleet typologian

suhdetta muihin kielentutkimuksen lähialoihin. Kielitypologia kehittää menetelmiä maailmanlaajuiseen kieltenvälisen variaation tutkimukseen. Sen perimmäiset tavoitteet liittyvät ihmiskielen luonteen selvittämiseen, ja se on tärkeä lähestymistapa yleisen kielitieteen keskeisiin kysymyksiin. Typologinen tutkimus luo uutta tietoa kielten eroista ja samankaltaisuuksista ja niihin vaikuttavista tekijöistä, joten se tarjoaa tärkeää tietoa selvitettäessä, miten kieli varioi, miten se muuttuu ja miten se on syntynyt.

**Aiheesta lisää:**

Aikhenvald, Alexandra & Dixon, R. M. W. (toim.). 2017. *The Cambridge Handbook of Linguistic Typology*. Cambridge: Cambridge University Press.

Comrie, Bernard. 1989. *Language Universals and Linguistic Typology*. (2. p.). Chicago (IL): The University of Chicago Press.

Croft, William. 2003. *Typology and Universals*. (2. p.). Cambridge: Cambridge University Press.

Dryer, Matthew S. & Haspelmath, Martin (toim.). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. [verkkoaineisto]. [viitattu 25.4.2018]. Saatavissa: https://wals.info.

Shopen, Timothy (toim.). 2007. *Language Typology and Syntactic Description*. (3 nidettä, 2. p.). Cambridge: Cambridge University Press.

Song, Jae Jung (toim.). 2010. *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press.

Velupillai, Viveka. 2012. *An Introduction to Linguistic Typology*. Amsterdam: John Benjamins.

## VIITTEET

1 Kiitämme artikkelimme toimittajia ja arvioijia hyödyllisistä kommenteista. Kaius Sinnemäki on saanut rahoitusta artikkelin kirjoittamiseen Suomen Akatemialta (hanke nro 296212) sekä Euroopan tutkimusneuvostolta (ERC) Euroopan unionin Horisontti 2020-tutkimus- ja innovaatio-ohjelmasta (sopimus nro 805371).

2 ”languages could differ from each other without limit and in unpredictable ways”.

3 *Tagmeemin* käsite syntyi strukturalismin parissa (Bloomfield 1933), mutta sen teki tunnetuksi Kenneth Pike (1912–2000), joka kehitti tagmeemianalyysista kattavamman kielen kuvauksen teorian (esim. Pike 1982). Lähestymistavan vaikeaselkoisuus johtuu osittain siitä, että siinä korostettiin kategorioiden kielikohtaisuutta siihen pisteeseen asti, että myös kategorioita kuvaaville termeille annettiin kielikohtaiset, esimerkiksi numeroista ja kirjaimista koostuvat nimet. Lisäksi kuvauksen taulukoiden ja formalismien tukena käytettiin usein vain hyvin niukalti suorasanaisia selityksiä.

4 Katso lisää Jörg Tiedemannin ylläpitämästä OPUS-verkkosivustosta (http://opus.lingfil.uu.se/), joka listaa avoimesti saatavilla olevia rinnakkaiskorpuksia.

5 Glossaussäännöt löytyvät osoitteesta https://www.eva.mpg.de/lingua/resources/glossing-rules.php. Esimerkissä käytetyt lyhenteet ovat: A = agentti, lauseen tekijä; P = patientti, teon kohde; CMPL = päättynyt aspekti; PST = menneen ajan tempus.

6 Kuva on tehty R-ohjelmalla (R Core Team 2018) käyttämällä sen paketteja ggplot2 (Wickham 2016), ggmap (Kahle & Wickham 2013) ja mapdata (Becker, Wilks & Brownrigg 2018).

## KIRJALLISUUS


Becker, Richard A., Wilks, Allan R. & Brownrigg, Ray. 2018. mapdata: Extra Map Databases. R package version 2.3.0. Saatavissa: https://CRAN.R-project.org/package=mapdata.

Bickel, Balthasar. 2007. Typology in the 21st century: Major current developments. *Linguistic Typology* 11:1, 239–251.

Bickel, Balthasar. 2015. Distributional typology: Statistical inquiries into the dynamics of linguistic diversity. Julkaisussa: Heine, Bernd & Narrog, Heike (toim.) *The Oxford Handbook of Linguistic Analysis.* (2. p.). Oxford: Oxford University Press, 901–923.

Bickel Balthasar, Witzlack-Makarevich, Alena, Choudhary, Kamal K., Schlesewsky, Matthias & Bornkessel-Schlesewsky, Ina. 2015. The neurophysiology of language processing shapes the evolution of grammar: Evidence from case marking. *PLoS ONE* 10:8, e0132819.

Bloomfield, Leonard. 1933. *Language*. New York (NY): Henry Holt.

Croft, William. 2003. *Typology and Universals.* (2. p.). Cambridge: Cambridge University Press.

Cysouw, Michael. 2005. Quantitative methods in typology. Julkaisussa: Altmann, Gabriel, Köhler, Reinhard & Piotrowski, Rajmund G. (toim.) *Quantitative Linguistics: An International Handbook*. Berlin: Mouton de Gruyter, 554–578.

Cysouw, Michael. 2010. Dealing with diversity: Towards an explanation of NP word order frequencies. *Linguistic Typology* 14:2, 253–287.

Cysouw, Michael & Wälchli, Bernhard (toim.). 2008. Parallel texts. Using translational equivalents in linguistic typology. *Language Typology and Universals (STUF)* 60:2, 95–181.

Dahl, Östen. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.

Dediu, Dan & Ladd, D. Robert. 2007. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences* 104:26, 10944–10949.

Derbyshire, Desmond C. 1979. *Hixkaryana*. Amsterdam: North-Holland.

Derbyshire, Desmond C. 1985. *Hixkaryana and Linguistic Typology*. Dallas (TX): Summer Institute of Linguistics.

Dixon, R. M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.

Dixon, R. M. W. 2009–2012. *Basic Linguistic Theory*. (3 nidettä). Cambridge: Cambridge University Press.

Dryer, Matthew. 1992. The Greenbergian word order correlations. *Language* 68:1, 81–138.

Dryer, Matthew. 1997. Are grammatical relations universal? Julkaisussa: Bybee, Joan Haiman, John & Thompson, Sandra (toim.) *Essays on Language Function and Language Type: Dedicated to T. Givon*. Amsterdam: John Benjamins, 115–143.

Dryer, Matthew. 2006. Descriptive theories, explanatory theories and Basic Linguistic Theory. Julkaisussa: Ameka, Felix, Dench, Alan & Evans, Nicholas (toim.) *Catching Language: The Standing Challenge of Grammar Writing*. Berlin: Mouton de Gruyter, 207–234.

Dryer, Matthew S. & Haspelmath, Martin (toim.). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. [verkkoaineisto]. [viitattu 25.4.2018]. Saatavissa: https://wals.info.

Eberhard, David M., Simons, Gary F. & Fennig, Charles D. (toim.). 2019. *The Ethnologue: Languages of the World*. (22. p.). Dallas (TX): SIL International. [verkkoaineisto]. [viitattu 16.12.2019]. Saatavissa: https://www.ethnologue.com.

Evans, Nick & Levinson, Stephen. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32:5, 429–448.

Everett, Caleb, Blasi, Damián E. & Roberts, Seán G. 2015. Climate, vocal folds, and tonal languages: Connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences* 112:5, 1322–1327.

Everett, Daniel. 2004. Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology* 46:4, 621–646.

Gil, David. 2013. Numeral classifiers. Julkaisussa: Dryer, Matthew & Haspelmath, Martin (toim.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. [verkkoaineisto]. [viitattu 25.4.2018]. Saatavissa: https://wals.info/chapter/55.

Graffi, Giorgi. 2010. The pioneers of linguistic typology: from Gabelentz to Greenberg. Julkaisussa: Song, Jae Jung (toim.) *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press, 25–42.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. Julkaisussa: Greenberg, Joseph H. (toim.) *Universals of Language*. Cambridge (MA): MIT Press, 73–113.

Greenberg, Joseph H. 1966. *Language Universals, with Special Reference to Feature Hierarchies*. The Hague: Mouton.
Greenberg, Joseph H. 1978. Diachrony, synchrony and language universals. Julkaisussa: Greenberg, Joseph H., Ferguson, Charles A. & Moravcsik, Edith A. (toim.) *Universals of Human Language I: Method and Theory*. Stanford (CA): Stanford University Press, 61–92.
Haiman, John. 1980. The iconicity of grammar: Isomorphism and motivation. *Language* 56:3, 515–540.
Hakulinen, Auli, Vilkuna, Maria, Korhonen, Riitta, Koivisto, Vesa, Heinonen, Tarja Riitta & Alho, Irja. 2004. *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Hammarström, Harald, Forkel, Robert & Haspelmath, Martin. 2019. *Glottolog 4.1*. Jena: Max Planck Institute for the Science of Human History. [verkkoaineisto]. [viitattu 1.2.2019]. Saatavissa: https://glottolog.org.
Haspelmath, Martin. 2005. Preface to the reprinted edition. Julkaisussa: Greenberg, Joseph H. *Language Universals, with Special Reference to Feature Hierarchies*. (Uudelleenjulkaistu p.). The Hague: Mouton, vii-xvii.
Haspelmath, Martin. 2007. Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11:1, 119–132.
Haspelmath, Martin. 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19:1, 1–33.
Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86:3, 663–687.
Haspelmath, Martin, Dryer, Matthew, Gil, David & Comrie, Bernard (toim.). 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.
Haugen, Einar. 1972. *Ecology of Language*. Stanford (CA): Stanford University Press.
Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
Himmelmann, Nikolaus P. 2000. Towards a typology of typologies. *Language Typology and Universals (STUF)* 53:1, 5–12.
Itkonen, Esa. 1991. *Universal History of Linguistics: India, China, Arabia, Europe*. Amsterdam: John Benjamins.
Itkonen, Esa. 1997. *Maailman kielten erilaisuus ja samuus*. Helsinki: Gaudeamus.
Jespersen, Otto. 1922. *Language: Its Nature, Development and Origin*. London: Allen & Unwin.
Joos, Martin. 1957. *Readings in Linguistics: The Development of Descriptive Linguistics in America since 1925*. Washington: American Council of Learned Societies.
Kahle, David & Wickham, Hadley. 2013. ggmap: Spatial Visualization with ggplot2. *The R Journal* 5:1, 144–161.
Karlsson, Fred. 2018. *Finnish: A Comprehensive Grammar*. London: Routledge.
Kittilä, Seppo. 2002. *Transitivity: Towards a Comprehensive Typology*. Turku: Turun yliopisto.
Lazard, Gilbert. 1992. Y a-t-il des categories interlangagières? Julkaisussa: Anschütz, Susanne & Heger, Klaus (toim.) *Texte, Sätze, Wörter und Moneme: Festschrift für Klaus Heger zum 65. Geburtstag*. Heidelberg: Heidelberger Orient-Verlag, 427–434.
Lupyan, Gary & Dale, Rick. 2010. Language structure is partly determined by social structure. *PLoS One* 5:1, e8559.

Martinet, André. 1955. *Économie des Changements Phonétiques: Traité de Phonologie Diachronique*. Berne: Francke.

Marttila, Annu. 2010. *A Cross-linguistic Study of Lexical Iconicity and Its Manifestation in Bird Names*. Väitöskirja. Helsinki: Helsingin yliopisto.

Maslova, Elena. 2000. A dynamic approach to the verification of distributional universals. *Linguistic Typology* 4:3, 307–333.

Miestamo, Matti. 2000. Towards a typology of standard negation. *Nordic Journal of Linguistics* 23:1, 65–88.

Miestamo, Matti. 2003. *Clausal Negation: A Typological Study*. Väitöskirja. Helsinki: Helsingin yliopisto.

Miestamo, Matti. 2005. *Standard Negation: The Negation of Declarative Verbal Main Clauses in a Typological Perspective*. Berlin: Mouton de Gruyter.

Miestamo, Matti. 2010. Joseph Greenberg ja kielitypologia. Julkaisussa: Haddington, Pentti & Sivonen, Jari (toim.) *Kielentutkimuksen modernit klassikot*. Helsinki: Gaudeamus, 153–182.

Miestamo, Matti. 2013. Kielten vertailu kielitypologisessa tutkimuksessa. Julkaisussa: Kolehmainen, Leena, Miestamo, Matti & Nordlund, Taru (toim.) *Kielten vertailun metodiikka*. Helsinki: Suomalaisen Kirjallisuuden Seura, 27–55.

Miestamo, Matti, Bakker, Dik & Arppe, Antti. 2016. Sampling for variety. *Linguistic Typology* 20:2, 233–296.

Mühlhäusler, Peter. 1996. *Linguistic Ecology: Language Change and Linguistic Imperialism in the Pacific Region*. London: Routledge.

Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago (IL): The University of Chicago Press.

Pajunen, Anneli (toim.). 1998. *Kieliopillistumisesta, analogiasta ja typologiasta*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Pajunen, Anneli (toim.). 2000. *Näkökulmia kielitypologiaan*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Perkins, Revere D. 1989. Statistical techniques for determining language sample size. *Studies in Language* 13:2, 293–315.

Piantadosi, Steven T. & Gibson, Edward. 2014. Quantitative standards for absolute linguistic universals. *Cognitive Science* 38, 736–756.

Pike, Kenneth L. 1982. *Linguistic Concepts: An Introduction to Tagmemics*. Lincoln (NB): University of Nebraska Press.

Ramat, Paolo. 2010. The (early) history of linguistic typology. Julkaisussa: Song, Jae Jung (toim.) *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press, 9–24.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Saatavissa: https://www.R-project.org/.

Rijkhoff, Jan & Bakker, Dik. 1998. Language sampling. *Linguistic Typology* 2:3, 263–314.

Salminen, Tapani. 1996. *Typological and Historical Studies in Language by Mikko Korhonen: A Memorial Volume Published on the 60th Anniversary of his Birth*. Helsinki: Suomalais-Ugrilainen Seura.

Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. Orlando: Harcourt Brace & co.

Shagal, Ksenia. 2019. *Participles: A Typological Study*. Berlin: De Gruyter Mouton.
Siewierska, Anna (toim.). 1998. *Constituent Order in the Languages of Europe*. Berlin: Mouton de Gruyter.
Sinnemäki, Kaius. 2008. Complexity trade-offs in core argument marking. Julkaisussa: Miestamo, Matti, Sinnemäki, Kaius & Karlsson, Fred (toim.) *Language Complexity: Typology, Contact, Change*. Amsterdam: John Benjamins, 67–88.
Sinnemäki, Kaius. 2011. *Language Universals and Linguistic Complexity: Three Case Studies in Core Argument Marking*. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-7259-8.
Sinnemäki, Kaius. 2014a. A typological perspective on differential object marking. *Linguistics* 52:2, 281–313.
Sinnemäki, Kaius. 2014b. Complexity trade-offs: A case study. Julkaisussa: Newmeyer, Frederick J. & Preston, Laurel B. (toim.) *Measuring Grammatical Complexity*. Oxford: Oxford University Press, 179–201.
Sinnemäki, Kaius. 2014c. Cognitive processing, language typology, and variation. *WIREs Cognitive Science* 5:4, 477–487.
Sinnemäki, Kaius & Di Garbo, Francesca. 2018. Language structures may adapt to the sociolinguistic environment, but it matters what and how you count: A typological study of verbal and nominal complexity. *Frontiers in Psychology* 9:1141. Saatavissa: https://doi.org/10.3389/fpsyg.2018.01141.
Stassen, Leon. 1985. *Comparison and Universal Grammar*. Oxford: Basil Blackwell.
Stassen, Leon. 2010. The problem of cross-linguistic identification. Julkaisussa: Song, Jae Jung (toim.) *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press, 90–99.
Suihkonen, Pirkko. 1990. *Korpustutkimus kielitypologiassa sovellettuna udmurttiin*. Helsinki: Suomalais-Ugrilainen Seura.
Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York (NY): Springer-Verlag.
Zipf, George K. 1949. *Human Behavior and the Principle of Least Effort*. Oxford: Addison-Wesley.

# Luku 15 Historiallinen ja vertaileva kielentutkimus

*Janne Saarikivi*

**Mikä?**

Historiallis-vertaileva kielentutkimus on kielitieteen vanhimpia ja laajimmin käytettyjä metodeja, ja sitä käytetään kaikkien kielikuntien tutkimuksessa. Sen avulla voidaan tutkia kielten välisiä historiallisia sukulaisuussuhteita ja lainakontakteja sekä kielten äänne- ja muotojärjestelmän ja sanaston historiaa. Tämä luku käsittelee kielitieteen historiallisia metodeja kiinnittäen päähuomion historiallis-vertailevaan eli komparatiiviseen metodiin. Historiallis-vertaileva metodi hyödyntää nykykielten tai kirjallisesti säilyneiden kielimuotojen aineistoja ja selittää niissä havaittavan variaation ja kielten väliset erot menneisyyden rekonstruoitujen vaihtelemattomien muotojen avulla.

**Katso myös:**

Luku 12 Kielikontaktien tutkimus

Luku 14 Kielitypologian menetelmät

Luku 16 Antropologinen lingvistiikka

Luku 18 Näkökulmia filologiseen tutkimukseen

Luku 19 Nimistöntutkimus

Luku 21 Sosiolingvistiikka

# 1. Johdanto: kielihistorian tutkimuskohteet

Kieli muuttuu ajan kuluessa. Historiallisen kielitieteen metodit mahdollistavat kielen muutoksen ja historian tutkimuksen pitkien aikojenkin kuluttua. Samalla ne tuottavat aineistoa, jota voidaan hyödyntää ihmisyhteisöjen alueiden, identiteettien, elinkeinojen sekä materiaalisen ja henkisen kulttuurin historian tutkimuksessa.

Historiallisen kielitieteen ytimessä on varsinainen kielihistoria, joka on kielen äännerakenteen, sanaston, muotorakenteen ja syntaksin historian tutkimusta. Myös kielellisen variaation tutkimuksessa on yleensä mukana kielen historia. Murteet eli dialektit, eri alueiden kielimuodot ja sosiolektit eli eri sosiaaliluokkien kielimuodot ovat historiassa syntyneitä. Variaationtutkimus ei kuitenkaan ole varsinaista kielihistorian tutkimusta, ellei siihen sisälly historiallista kysymyksenasettelua (ks. Sosiolingvistiikka tässä kirjassa).

Historiallisen kielitieteen metodit kohdistuvat sekä kielen sisäiseen että kielen ulkoiseen historiaan. Sisäisen kielihistorian tutkimuksella tarkoitetaan kielen sanaston ja kieliopin eli taivutusmuotojen ja rakenteiden muutoksen tutkimusta ja niiden varhaisempien muotojen rekonstruoimista. Ulkoista kielihistoriaa taas ovat kielten puhuma-alueiden, kieliä ylläpitävien sosiaalisten verkostojen ja uusia vaikutteita tuovien kulttuurikontaktien historia. Ulkoinen kielihistoria pyrkii usein monitieteiseen lähestymistapaan. Kielellisten materiaalien rinnalla keskeisiä ovat tällöin mm. arkeologian, genetiikan, uskonto- ja kansatieteen aineistot. Silti myös ulkoinen kielihistoria pohjautuu aina sisäisen kielihistorian luomaan kuvaan kunkin kielen edeltäjistä, sukukielistä ja kontakteista.

Historiallinen kielitiede käyttää monia eri tutkimusmenetelmiä. Filologiset metodit ovat vanhojen tekstien tutkimusta, erityisesti tekstien varhaisimman muodon rekonstruoimista eri käsikirjoitusvariantteja vertailemalla. Samassa yhteydessä tutkitaan sanojen ja käsitteiden merkityksenmuutosta, johto-opillista kehitystä ja muuta variaatiota sekä näiden heijastamia kulttuurisia muutoksia. (ks. Näkökulmia filologiseen tutkimukseen tk.)

Historiallinen sosiolingvistiikka rekonstruoi menneisyyden kielenkäytön variaatiota, mm. yhteisöjen monikielisyyttä, kielellisiä verkostoja ja kontakteja sekä kieliyhteisöjen kulttuuria (ks. Sosiolingvistiikka tk.). Historiallisen nimistöntutkimuksen kohteena on paikan- ja henkilönimien alkuperä sekä yhteisöt, jotka ovat käyttäneet ja levittäneet nimiä. Nimistöntutkimuksen avulla on mahdollisuus selvittää eri alueiden aikaisempia kielioloja, esim. alueilla aiemmin puhuttuja kieliä, sekä varhaisia väestöliikkeitä ja kielenvaihtoja (ks. Nimistöntutkimus tk.).

Kontaktilingvistiikka jäljittää kielten ja kieliyhteisöjen historiallisia kontakteja ja kielen muuttumista kontaktien myötä. Laajassa mielessä kontaktilingvistiikkaan kuuluu kaikki kontaktin aiheuttama kielen muutos ja sen erottaminen muista syistä aiheutuneista muutoksista. Suppeassa mielessä (laina)sanaston alkuperän ja siinä ilmenevien kontaktien tutkimus on usein etymologian tutkimuskohde ja varsinainen kontaktilingvistiikka selvittää kielellisten rakenteiden muutosta esimerkiksi morfosyntaksin ja äännejärjestelmän alalla (ks. Kielikontaktien tutkimus tk.).

Tämä luku esittelee kuitenkin pääasiassa historiallis-vertailevaa metodia, joka on varsinaisen kielihistorian ydin. Historiallis-vertaileva metodi tuottaa päätuloksenaan kielisukulaisuuteen eli kielenaineksen periytyvyyteen, muutokseen ja lainaamiseen liittyvää tietoa. Tästä johtuen se on avainasemassa myös tutkittaessa historiallisten kielikontaktien vaikutuksia eli niitä asioita, jotka tavalla tai toisella ovat lainautuneet muista kielistä; nimittäin jotta kielikontaktin vaikutukset voidaan tunnistaa, on tiedettävä, mikä kielessä on aikaisemmista kielivaiheista periytyvää. Historiallis-vertailevan metodin tutkimuskohteita ovat etymologia eli sanaston historia, historiallinen fonologia eli äännehistoria, historiallinen morfologia eli muotohistoria sekä historiallinen syntaksi eli historiallinen lauseoppi ja kielen rakenteiden muutoksen tutkimus.

# 2. Miksi kieli muuttuu?

## 2.1. Äänteenmuutos

Kielen muutokseen on monia syitä. Monet äänteenmuutokset liittyvät ääntämisen yksinkertaistumiseen. Ne voivat koskea sekä kielen foneettista järjestelmää eli ääntämistapoja että fonemaattista järjestelmää eli äännejärjestelmää. Historiallis-vertaileva metodi on kiinnostunut lähinnä jälkimmäistä tyyppiä edustavista muutoksista. Tällaisia ovat esimerkiksi äänteiden assimilaatiot naapuriäänteisin, äänteiden katoamiset ja uusien äänteiden syntymiset. Muutosten taustalla on usein pyrkimys puhua vähemmän ääntöenergiaa käyttäen. Käytännössä tällaiset ilmiöt kohdistuvat kielissä samantyyppisiin äännekonteksteihin, mikä mahdollistaa niiden etsimisen historiassa. Esimerkiksi klusiilien muuttuminen affrikaatoiksi ja frikatiiveiksi etuvokaalin edellä (*ki* > či > *si*, *ti* > či > *si*, *pi* > *pf* > *f*) tai painottoman vokaalin katoamiset ovat muutoksia, joita tunnetaan lukuisista eri kielistä ja eri yhteyksistä.

## 2.2. Analogia

Äänteenmuutokset ovat yleensä säännöllisiä eli koskevat kaikkia samanlaisia äännesekvenssejä kielessä. Kielijärjestelmän tasolla ne kuitenkin aiheuttavat epäsäännöllisyyttä. Analogia on kielen epäsäännöllisyyksien mukauttamista tai korjaamista tavalla tai toisella säännönmukaiseksi. Toisin kuin äännemuutokset analogiset muutokset ovat yleensä epäsäännöllisiä, eli ne tapahtuvat vain yhden sanan tai muodon kohdalla kerrallaan. Toisaalta kielijärjestelmän tasolla ne lisäävät kielen säännöllisyyttä.

Esimerkiksi suomen sanasto kuuluu erilaisiin taivutustyyppeihin, joista osa on syntynyt säännöllisen äännekehityksen myötä, toiset taas siten, että taivutusmuotoa on yleistetty uusiin ympäristöihin, koska se on läpinäkyvä ja helposti segmentoitava. Turun murteelle tyypillinen *si*-imperfekti *menis, tulis* 'meni, tuli' on syntynyt yleistämällä ns. supistumaverbien (*ruveta* : *rupesi*, *karata* : *karkasi*) imperfektin *si*-pääte muihin

taivutustyyppeihin. Virossa *si*-imperfektin voi muodostaa useimmista verbeistä. Se yksinkertaistaa huomattavasti monien eri verbityyppien taivutusta esim. suomeen verrattuna.

Monesti analogia tasoittaa kielessä tapahtuneita äänteenmuutoksia. Esimerkiksi suomen sanan *lahti* aikaisempi asu on ollut *laksi*, joka on syntynyt vielä varhaisemmasta asusta **lakti*. Tämä on osallistunut erilaisiin säännönmukaisiin äänteenmuutoksiin. Lopulta sanan nominatiivimuoto *laksi* ei enää muistuttanut kovin paljon muiden sijojen muotoja (*lahden*, *lahdessa* jne.). Vaikka samanlainen paradigma on ollut myös esim. sanassa *kaksi* (*kahden*, *kahdessa*), on *lahti*-sanan kohdalla jostain syystä koettu tarpeelliseksi ottaa uusi analoginen nominatiivimuoto *lahti* käyttöön äänteellisesti säännönmukaisen *laksi*-muodon asemesta. Analogian mallina ovat toimineet muilla tavoin syntyneet äänteellisesti läheiset sanat, esim. riimisana *lehti* (vrt. Lehtinen 2007, 98–99).

## 2.3. Sosiaalisesti ja kulttuurisesti muuttuva kieli

Kaikki kielenmuutos ei johdu ääntämistapojen muutoksesta. Kielenmuutoksessa on kyse aina myös sosiaalisesta toiminnasta. Puhe ja kielenkäytön tyyli ovat tärkeitä asioita sekä yksilön persoonan että yhteisön identiteetin luomisessa, ylläpitämisessä ja muutoksessa. Yleensä kaikki kielenkäyttäjät pyrkivät puhumaan hieman toisistaan poiketen mutta kopioivat silti toistensa antamia malleja. Näin kielenkäytön muodit muuttuvat vähän kerrallaan; yksi sana jää pois käytöstä ja toinen, muodikkaampi sana ottaa sen paikan. Esimerkiksi *hullujenhuone* on korvaantunut ensin sanalla *mielisairaala*, joka sittemmin on korvautunut ilmauksella *psykiatrinen sairaala*. Muutos kuvastaa yhteiskunnan arvojen ja psykiatrisiin sairauksiin liittyvien käsitysten muutosta. Tyypillistä on myös, että tietystä kontekstista tuttua idiomia tai rakennetta aletaan yleistää uusiin yhteyksiin. Aluksi sanotaan esimerkiksi, että *ei ole kaikki kotona*, ja pian tämän mallin mukaan, että *ei ole kaikki kynät penaalissa* tai *ei ole kaikki muumit laaksossa* (Kortelainen 2012).

Kielijärjestelmän muutoksessa on perinteisesti katsottu olevan oleellista muutoksen sosiaalisen houkuttelevuuden eli prestiisin. Esimerkiksi

Suomen nykyisessä kielitilanteessa englanninkieliset ilmaukset leviävät erityisesti nuorisokielessä ja monenlaisissa ammattikielissä, koska ne liittyvät houkuttelevina pidettyihin sosiaalisiin roolimalleihin.

Kielellinen kulttuuri, mm. käsitykset normeista sekä ei-toivottavista kielenkäytön tavoista, on omiaan vaikuttamaan käyttämäämme kieleen (Schiffman 1996; Toivanen & Saarikivi 2016). Tähän liittyvät myös kysymykset enemmistö- ja vähemmistökielten välisistä suhteista ja siitä, missä sosiaalisissa verkostoissa ja tilanteissa vähemmistökielten käyttäminen on kulloinkin mahdollista. Eri maiden ja alueiden kielelliset kulttuurit ovat tässä suhteessa hyvin erilaisia; yhdet ovat tolerantteja suhteessa monikielisyyteen ja murteelliseen puheeseen, toisissa vallitsee normitetun valtakielen suosiminen.

## 2.4. Lainautuminen

Kielenmuutos on usein sidoksissa kulttuuriympäristön muutokseen. Näin on erityisesti lainautumisen tapauksessa. Kulttuuriympäristön muuttuessa saadaan naapurikielistä uusia sanoja ja ilmauksia nimityksiksi asioille, jotka ovat aikaisemmin olleet tuntemattomia. Itse asiassa suomen johtamattomien ns. perussanojen valtaosa on lainasanoja.

Lainautuminen voi kohdistua paitsi sanastoon myös kielen äännejärjestelmään ja lauserakenteeseen. Tyypillistä on, että intensiivisen, yksilöiden kaksikielisyyttä sisältävän kielikontaktin tapauksessa kielestä toiseen kopioituu myös idiomeja ja rakenteita (engl. *in the long run* → *pitkässä juoksussa*) sekä ääntämistapoja ja äänteitä. Vastaavasti suomen kirjakieleen kuuluvat äänteet *b*, *d*, *g* ja *f* esiintyivät alkuaan lähinnä lainasanastossa (*banaani*, *dollari*, *fraasi*), mutta niiden yleistyttyä ne ovat eräissä suomen varieteeteissa alkaneet esiintyä myös muussa sanastossa (esim. Stadin slangin *galsa* 'kylmä' < *kalsa* < ruotsi *kall*, *dallaa* < *tallata*).

Vaikka intensiivinen sosiaalinen kontakti johtaa lainaamisen lisääntymiseen, on toisaalta osoitettu, että kielenmuutosten kannalta oleellisia ovat usein pikemminkin heikot kuin vahvat sosiaaliset linkit (Milroy 1992). Vahvat linkit taas pitävät huolta kielijärjestelmän tietystä stabiilisuudesta ja riittävästä muuttumattomuudesta. Uusia kielenpiirteitä käyttäessään

ihmiset eivät kopioi niinkään oman perheensä tai lähiympäristönsä kieltä vaan toistavat muualta kuulemiaan sanoja ja rakenteita.

## 2.5. Johtaminen

Uusia sanoja nimityksiksi uusille asioille saadaan paitsi lainaamalla myös johtamalla ja yhdistämällä vanhaa sanastoa. Suomen kielessä yhdistämis- ja johtomahdollisuudet ovat hyvin monipuoliset. Monissa nykysuomen sanoissa on lukuisia peräkkäisiä sanoja ja johdinaineksia. Näin ovat viime vuosikymmeninä syntyneet mm. uudissanat *elämyspuisto* (viime kädessä sanoista *elää* ja *puu*), *asiakaspalvelu* (sanoista *asia* ja *palvoa*), *sote-uudistus* (sanoista *sosiaalinen, terve* ja *uusi*) ja lukemattomat muut. Esimerkiksi tässä artikkelissa yli puolet sanoista lienee yhdistelmiä ja johdoksia muista sanoista. Suuri osa tällaisesta sanastosta on melko uutta, vaikka sen rakenneosaset ovat usein ikivanhoja.

## 2.6. Kielenvaihto

Myös kielenvaihto on eräänlainen kielenmuutos, vaikkakin se johtaa kielen katoamiseen ainakin osassa sen puhujayhteisöä. Kielenvaihtojen mekanismit ovat yleensä kombinaatio pakkoa ja vapaaehtoisuutta. Tavallisesti kyseessä on tilanne, jossa jokin kieli kadottaa sosiaalisen prestiisinsä valtasuhteiden muutoksen tai yhteisön hajoamisen seurauksena.

Kielenvaihdoista on tullut tavallisia modernissa maailmassa, jossa dominoivat puhutun kielen asemesta kirjakielet. Niihin liittyy kielellinen standardi, joka on opittava vuosia kestävässä koulutusprosessissa. Niitä käytetään myös työvälineenä moderneissa ammateissa, kuten opettajien, juristien, suunnittelijoiden ja konsulttien töissä, joissa vähemmistökielet ovat tarpeettomia. Modernia yhteiskuntaa luonnehtiikin, että yhä suurempi osa työelämästä on nimenomaan kirjoitetun kielen lukemista ja tuottamista.

Kielenvaihtoja on silti ollut runsaasti myös esihistoriallisena aikana. Monet esihistorian kieliyhteisöt ovat todennäköisesti suhtautuneet

kieliin melko pragmaattisesti kuten eräät metsästäjä-keräilijä-yhteisöt nykyäänkin. Niissä monikielisyys on ollut hyvin yleistä. Esimerkiksi Länsi-Siperiassa ihmiset ovat saattaneet osata kolmea neljää, jopa viittä kieltä alueesta riippuen, samoin historiallisella Inkerinmaalla. Vastaavaa monikielisyyttä on edelleen säilynyt mm. Afrikassa (Lüpke 2013).

Monikielisissä yhteisöissä on helposti käynyt siten, että jokin tietty kieli on päässyt voitolle – yleensä koska kyseessä on ollut sosiaalisesti, kulttuurisesti tai taloudellisesti vahvimman ryhmän kieli – ja hävittänyt altaan kielellistä monimuotoisuutta. Esimerkiksi nykyisen Ranskan alueella puhutaan romaanista kieltä, mutta antiikin aikana alue oli kielellisesti kelttiläinen. Tiedetään, että kelttiläinenkään kielimuoto ei ole alueella alkuperäinen, vaan tätä ovat edeltäneet tuntemattomat paleoeurooppalaiset kielimuodot. Niistä osa on mahdollisesti ollut sukua baskille, osa taas kokonaan tuntemattomia. Vastaavasti Anatolian eli nykyisen Turkin alueella, jossa turkki on nyt valtakieli, puhuttiin vielä keskiaikana kreikkaa, jota taas edelsi valtakielenä persian kieli, jota edelsi heetin kieli. Britannian alueella tiedetään englannin esimuotojen syrjäyttäneen mm. kelttiläisiä ja skandinaavisia kieliä, mutta nämäkään eivät edusta alueen varhaisimpia kieliä, jotka ovat todennäköisesti kuuluneet kokonaan kadonneisiin kielikuntiin. Suomen alueella on esihistorian ja historian aikana tapahtunut ainakin kaksi lähes täydellistä kielenvaihtoa, ensin tuntemattomista muinaiskielistä saameen ja sitten saamesta suomeen, osin historiallisena aikana. Eräillä alueilla on suomalaisväestö vaihtanut kielensä vielä ruotsiin, ja tätäkin myöhemmin on taas ruotsinkielinen väestö siirtynyt jälleen suomeen (Suomen osalta ks. Aikio 2012; Saarikivi 2006; ks. myös Sosiolingvistiikka tk.).

Jos jotain kieltä lakataan puhumasta kaikissa yhteisöissä, seurauksena on kielikuolema. Kielikuolemat ovat tällä hetkellä yleisiä. Arvioidaan, että jopa 50–90 % maailman kielistä voi kuolla muutaman sukupolven aikana. Historiallinen kielitiede voi jossain määrin tutkia sekä alueellisesti että kokonaan kuolleita kieliä, jos ne ovat jättäneet substraattivaikutuksia nykykieliin (ks. alla).

# 3. Historiallisen kielitieteen tutkimuskohteet

## 3.1. Historiallis-vertailevan metodin tutkimuskohteet

Sisäinen ja ulkoinen kielihistoria jakautuvat moniin tutkimussuuntiin. Perustan molemmille muodostaa historiallis-vertaileva eli historiallis-komparatiivinen metodi, joka on jatkuvasti täydellistynyt ja täsmentynyt (vrt. esim. Paul 1881; Ravila 1946; Anttila 1988; Fox 1995; Campbell 1998.

Historiallis-vertailevan metodin ytimessä ovat sanojen ja kielen affiksien (eli suffiksien, prefiksien ja muiden merkitystä kantavien sidonnaisten morfeemien) alkuperän ja äänneasun selvittäminen. Etymologia selvittää kielen sanaston ja kieliopillisten affiksien alkuperää sekä kielen sanastokerrostumia. Siihen on läheisessä yhteydessä myös historiallinen semantiikka eli merkitysmuutosten ja käsitteiden historian selvittäminen. Historiallinen fonologia taas selvittää kielen äännerakenteen historiaa ja siinä tapahtuneita muutoksia. Se on vuorovaikutussuhteessa etymologiaan, joka tuottaa sille aineistoa. Voi sanoa, että etymologia ja historiallinen fonologia edellyttävät toisiaan kuten kruuna ja klaava tai vasen ja oikea.

Historiallis-vertailevan metodin hallitseva kielitieteilijä osaa arvioida sanojen ikää ja alkuperää niiden äänteistön ja merkityksen avulla, selittää sanoissa havaittavia äännevaihteluja historiallisesti ja kykenee näkemään, millaisista varhaisemmista muodoista sanat ovat saattaneet kehittyä. Tässä työssä apuna on tieto kielen fonotaksista eli kielen sanojen fonologisesti mahdollisista sanahahmoista. Esimerkiksi suomen kielessä ei alkuaan ole ollut sanoja, joissa on sananalkuisia konsonanttiyhtymiä. Niinpä sellaiset sanat kuin *trooli*, *kranttu* tai *strategia* ovat kaikki nuoria lainoja naapurikielistä.

Historiallinen morfologia selvittää kielen taivutusmuotojen historiaa ja muutosta. Koska monet taivutusmuodot ovat alkuaan itsenäisiä sanamuotoja, ei etymologian ja historiallisen morfologian välinen raja ole kovinkaan selkeä. Kieliopillistuminen eli grammatikalisaatio on prosessi, jossa itsenäiset sanat ja niiden taivutusmuodot muuttuvat kieliopillisiksi muodoiksi, mm. morfeemeiksi. Kieliopillistumista pidetään

lähinnä tutkimushistoriallisista syistä omana tutkimusalanaan, mutta se on kiinteästi sidoksissa historiallis-vertailevaan metodiin.

Vaikka historiallis-vertailevan metodin aineistona on konkreettinen kielenaines, sanojen ja affiksien fonologinen hahmo, se selittää nykykielten vaihtelun menneisyyden abstraktioilla eli kantakielen rekonstruktioilla. Yleensä kantakielet ovat säännönmukaisempia kuin nykykielet ja niiden morfologinen variaatio on vähäisempää. Myös muinaisissa kielimuodoissa on silti varmasti ollut variaatiota, mutta tämä variaatio ei välttämättä käy ilmi rekonstruktioista.

Ajatus kantakielten eriytymisestä tytärkieliin saa mallinsa todellisista kielenmuutostapauksista. Esimerkiksi latina on muuttunut nykyisiksi romaanisiksi kieliksi, sanskrit nykyisiksi Etu-Intian kieliksi jne. Historiallis-vertaileva metodi pohjautuu oletukseen, että vastaava prosessi on tapahtunut useimpien muidenkin tai jopa kaikkien kielten historiassa. Tulee kuitenkin huomata, että nykyisten romaanisten kielten pohjalta rekonstruoitu latina saattaisi monilta osin erota todellisuudessa olemassa olleesta latinan kielestä.

Suurin osa maailman kantakielistä ei ole kirjallisesti säilynyt kuten latina, vaan ne tunnetaan ainoastaan rekonstruktioina. Kun historiallis-vertaileva menetelmä rakentaa nykykieliä vertailemalla kantakielten rekonstruktion, se luo samalla kantakielten taksonomian eli periytyvyysjärjestelmän. Esimerkiksi suomen kieli ja viron kieli yhdessä monien pienten itämerensuomalaisten kielten kanssa polveutuvat kantasuomesta, joka puolestaan polveutuu kantauralista. Kantauralista polveutuu kantasuomen lisäksi kuudesta kahdeksaan muuta kantakieltä, jotka ovat edelleen jakautuneet kymmeniksi nykykieliksi. Vastaavasti indoeurooppalaisen kantakielen tytärkieliä ovat latina, sanskrit sekä rekonstruoidut kantagermaani, kantaslaavi, kantakeltti ym. Nämä jakaantuvat sadoiksi nykykieliksi: kantagermaanista on tullut englantia, ruotsia, saksaa ym., kantaslaavista venäjää, puolaa, serbiä, jne.

Yhdessä kaukaisimpaan yhteiseen rekonstruoitavaan kantakieleen palautuvat kielet muodostavat kielikunnan. Kielikuntia tunnetaan maailmassa n. 250. Historiallis-vertailevaa menetelmää voi soveltaa minkä hyvänsä kielikunnan tutkimukseen, mutta käytännössä monienkaan kielikuntien historiallis-vertaileva tutkimus ei ole kovin edistynyttä.

Esimerkiksi sellaisista maailmankielistä kuin nykykiinasta tai arabiasta ei ole ilmestynyt edes varsinaista etymologista sanakirjaa. Tämä johtuu monien kielten huonosta dokumentoinnista ja eri kielikuntien hyvin eriasteisesta aikasyvyydestä. Historiallis-vertaileva metodi toimii parhaiten suhteellisen nuorten kielikuntien tutkimuksessa.

Historiallis-vertaileva menetelmä näyttää luokitelleen maailman kielikunnat suhteellisen kattavasti. 1950-luvun jälkeen on löytynyt vain muutamia uusia kielikuntia, vaikka kielten dokumentaatio on jatkuvasti edennyt. Julkisuuteen silloin tällöin päätyvät sensaatiomaiset uutiset, joiden mukaan suomen kieli on sukua esimerkiksi sumerille tai baskille, ovat yleensä historiallis-vertailevan standardimetodin kannalta perusteettomia uutisankkoja.

Nykyisin tiedetään myös, että maailmassa on myös eräitä alueita, joilla ei ylipäätään esiinny suuria kielikuntia. Näitä ovat erityisesti metsästäjä-keräilijäväestöjen asuttamat alueet, kuten Papua tai Australia. Tällaisilla alueilla kielikunnat koostuvat tyypillisesti melko pienestä määrästä kieliä. Mahdollisesti syynä tähän ilmiöön on se, että metsästäjä-keräilijöiden pienet yhteisöt eivät enimmäkseen mahdollista huomattavaa väestönkasvua, mikä useissa tilanteissa on kielialueen laajenemisen edellytys. Mikäli näin on, suuret kielikunnat ovat ihmiskunnan historiassa suhteellisen nuori ilmiö ja liittyvät vasta maanviljelyksen syntyyn ja sen mahdollistamaan väestöjen moninkertaistumiseen.

## 3.2. Historiallis-vertailevan metodin historiaa

Ihmiset ovat aina olleet tietoisia lähisukuisten kielten välisistä samankaltaisuuksista. Kun Euroopan kielistä suuri osa polveutuu suoraan latinasta, jota käytettiin ainoana kirjakielenä satojen vuosien ajan, ajatus kielten polveutumisesta toisistaan on aina ollut luonteva. Kukaan ei liene koskaan tieteellisesti todistanut, että esimerkiksi ranska tai italia polveutuu latinasta. Asia on ollut itsestään selvä osa kieliyhteisön tietoisuutta. Myöskään suomen ja viron välistä sukulaissuhdetta ei ole koskaan todistettu missään toisin kuin esimerkiksi suomen ja saamen tai unkarin välinen sukulaisuus, joka voidaan havaita vain tieteellisten metodien avulla.

Historiallis-vertaileva metodi on alkuaan syntynyt renessanssiaikana harjoitettujen kielivertailujen pohjalle. Renessanssioppineet pyrkivät selvittämään maailman kaikkien kielten alkuperän ja suhteen Raamatun pyhiin kieliin, kreikkaan ja hepreaan. Samalla kerättiin aineistoja maailman kielistä. Ensimmäisinä nykyisenkaltaisen historiallis-vertailevan metodin edustajina mainitaan usein unkarilainen jesuiittamunkki János Sajnóvics (Sajnóvics 1964), joka todisti vuonna 1770 suomen, saamen ja unkarin välisen yhteyden, sekä brittiläinen sir William Jones, joka vuonna 1776 pitämässään esitelmässä todisti latinan, kreikan ja sanskritin kielten yhteisen alkuperän. Molemmat näistä todistuksista sisälsivät sanaston ja kieliopillisten päätteiden vertailua sekä oletuksen, jonka mukaan samankaltaiset nykykielet ovat kehittyneet yhdestä ja samasta kadonneesta muinaiskielestä.

1800-luvulla tämänkaltaiset melko epätäsmälliset ja jäsentymättömät vertailut kasvoivat systemaattiseksi tieteeksi. Historiallis-vertailevan metodin varhaisina pioneereina mainitaan tanskalainen Rasmus Rask, saksalainen Franz Bopp sekä myös kansansatujen kokoajana tunnettu Jacob Grimm, joka tunnetaan ns. Grimmin laista. Se selittää mm. latinan ja germaanisten kielten klusiilien ja frikatiivien välisen suhteen eli vaikkapa sanaparien latinan *pisces* ~ englannin *fish*, latinan *pos* : *podus* ~ englannin *foot* välisen suhteen olettamalla, että germaanisissa kielissä on tapahtunut klusiilien spirantistuminen (**p* > *f* jne.) tietyissä asemissa. Samalla se selittää myös kantaindoeuroopan soinnillisten ja aspiroituneiden klusiilien muutokset ja nykyedustuksen germaanisissa kielissä.

1800-luvun loppuun mennessä historiallis-vertaileva menetelmä kehittyi ns. nuorgrammatiikaksi, joka perustui ajatukseen kielenmuutoksen tietynlaisesta mekanistisuudesta. Nuorgrammatiikan ydinolettama oli, että kaikki kielenmuutos perustuu epäsäännöllisyyttä aiheuttaviin äänteenmuutoksiin tai säännöllisyyttä palauttavaan analogiaan. Samoin oletettiin, että kaikki äänteenmuutokset pohjautuvat lähtökohdiltaan poikkeuksettomiin äännelakeihin. Tämän käsityksen kielihistoriasta esitti Hermann Paul vuonna 1881 ilmestyneessä teoksessaan *Prinzipien der Sprachgeschichte* (Kielihistorian perusteet), joka on edelleen arvostettu klassikkoteos.

1800-luvun loppu ja 1900-luvun alkupuoli olivat historiallis-vertailevan kielitieteen kannalta ns. normaaliparadigman eli saman-

kaltaisia tutkimusasetelmia toistavan ja tietoa kartuttavan mutta metodisesti konservatiivisen tutkimuksen aikaa. Useissa Euroopan maissa ilmestyi eri murteiden ja kielten äännehistorioita ja lainasanakerrostumia käsitteleviä tutkimuksia. Varhaisia kirjoitettuja lähteitä käsittävän indoeurooppalaisen kielikunnan tapauksessa fokus oli kantakielten rekonstruoinnissa. Uralilainen historiallis-vertaileva kielentutkimus taas kehittyi ennen kaikkea indoeurooppalaisista kielistä saatujen lainakerrostumien tutkimuksena. Nykypäivänäkin on yhä tilanne, että indoeurooppalainen kantakieli on paljon tarkemmin rekonstruoitu kuin uralilainen kantakieli. Vastaavasti uralilaisten kielten monet vanhatkin lainasanakerrostumat on usein kohtalaisen hyvin kuvattu, kun taas kantakielen rekonstruktiot elävät koko ajan ja niihin liittyy paljon epävarmuustekijöitä.

Jo varhain ymmärrettiin, että poikkeuksettomat äännelait ja analogia eivät voi selittää kaikkea kielen muutosta. Erityisesti murteiden tutkimus osoitti, että sanaston kehitykseen liittyy myös kontaminaatio ja kansanetymologia (ks. alla). Kun Hermann Paulin ja nuorgrammatiikan aksiooma oli *Die Lautgesetze kennen keine Ausnahme* (’äännelait ovat poikkeuksettomia’), alettiin tämän rinnalla esittää murremaantieteen ja etymologian aksioomaa *Chaque mot a son histoire* (’jokaisella sanalla on itsenäinen historiansa’). Jälkimmäinen on pantu ranskalaisen kielihistorioitsija Antoine Meillet’n nimiin.

# 4. Historiallis-vertailevan metodin soveltamisesta

## 4.1. Peruskäsitteistä

Keskeisiä historiallis-vertailevan metodin käsitteitä ovat *äännevastaavuus, vastine, äännelaki, sisäinen ja ulkoinen rekonstruktio, periytyminen, lainautuminen* ja *substraatti*.

Kahden eri kielen historiallisesti toisiaan vastaavia äännekonteksteja kutsutaan äännevastaavuuksiksi ja toisiaan vastaavia sanoja *vastineiksi*

eli *kognaateiksi*. Esimerkiksi suomen sanan *käsi* vastineita ovat mm. viron *käsi*, pohjoissaamen *giehta*, niittymarin *kid*, komin *ki* ja unkarin *kéz*. Saamen *giehta*-sanaa suomeen verrattaessa voidaan havaita äännevastaavuudet 1. tavun ä ~ *ie*, 2. tavun *i* ~ *a*, sananalkuinen *k* ~ *g* ja niin edelleen. Säännölliset äännevastaavuudet voidaan yleensä selittää ns. äännelakien avulla. Äännelaki on samassa ympäristössä samalla tavoin tapahtunut äänteenmuutos. Esimerkiksi suomen sanan *käsi* ja saamen sanan *giehta* välisen äännevastaavuuden selittää muun muassa suomen kielessä tapahtunut *ti* > *si* -muutos. Kyseessä on kaikkia *i*:n etisiä *t*-äänteitä koskenut äännelaki, jonka mukaan ne ovat muuttuneet *s*-äänteiksi. Tämän lisäksi sanassa on tapahtunut mm. *t*-äänteen soinnillistuminen umpitavun etisessä asemessa, mikä on tuottanut ns. astevaihtelun eli *d*- ja *t*-äänteiden vaihtelun, joka näkyy muodoissa *käden* ~ *kättä*. Samanlainen vaihtelu on myös esim. sanoissa *tosi*, *vesi*, *mesi* jne.

Äännelait voivat olla kontekstiehtoisia tai kaikissa konteksteissa tapahtuneita. Esim. *ti* > *si* -muutos on kontekstiehtoinen äännelaki, joka koskee *t*-äännettä *i*:n edellä. Muissa äänneympäristöissä *t* säilyy *t*-äänteenä jopa saman sanan taivutusparadigmassa (vrt. esimerkiksi *käsi* : *kätenä*). Vastaavasti toinen kantasuomessa tapahtunut äänteenmuutos, ns. *š* > *h* -muutos on kaikkia konteksteja koskeva. Kaikissa konteksteissa tapahtuneet äännelait ovat huomattavasti harvinaisempia kuin kontekstiehtoiset. Kontekstiehtoisiin kehityksiin kuuluu esimerkiksi se, että painolliset ja painottomat tai sananalkuiset ja sanansisäiset äänteet käyttäytyvät kielen historiassa eri tavoin. Tavallista on niin ikään, että äänteet assimiloituvat eli mukautuvat viereisiin äänteisiin tai että ensimmäisen tavun ja toisen tavun vokaali vaikuttavat toisiinsa kuten suomen kielen vokaalisoinnussa, jossa samassa sanassa ei voi esiintyä etu- ja takavokaaleja tietyin ehdoin.

Käytännössä on havaittu, että äännelaillinen kehitys ohjaa kielen muutosta hyvin erilaisissa kielissä. Esimerkiksi Grimmin lakia (ks. edellä) vastaava suhde vallitsee myös suomen ja unkarin välillä. Suomen *pää*-sanaa vastaa unkarissa *fej*, *puu*-sanaa *fa* ja *poika*-sanaa *fiú*. Myös saksan kirjakielessä (tai ns. yläsaksalaisissa murteissa) on tapahtunut sananalkuisen *p*:n affrikoituminen eli muuttuminen hälyäänteeksi etuvokaalin edellä, mikä ilmenee verrattaessa ruotsin sanoja *penning* ja

*peppar* sekä saksan sanoja *pfennig* ja *pfeffer*. Maailman kielissä tapahtuneet äänteenmuutokset toistavat siis yleensä samoja muutostyyppejä.

## 4.2. Sisäinen rekonstruktio

Päävälineinä historiallis-vertailevan metodin soveltamisessa ovat sisäinen rekonstruktio ja ulkoinen rekonstruktio. Seuraavassa tarkastellaan molempia hieman yksityiskohtaisemmin.

Sisäinen rekonstruktio käyttää aineistonaan yhden kielen sisäisiä äännevaihteluita. Tätä voidaan havainnollistaa vertaamalla esimerkiksi eri taivutustyyppeihin kuuluvien suomen kielen nominien taivutusta:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *lasi:* | *lasin:* | *lasia* | *kivi:* | *kiven:* | *kiveä* | *käsi:* | *käden:* | *kättä* |
| *laki:* | *lain:* | *lakia* | *laki:* | *laen:* | *lakea* | *tosi:* | *toden:* | *totta* |

Aineistosta voimme havaita, että suomen kielessä on kahdenlaisia sananloppuisia *i*-sekvenssejä. Toiset niistä ovat vaihtelemattomia, mutta toiset vaihtelevat sanansisäisessä asemassa esiintyvän *e*-sekvenssin kanssa. Tästä on päätelty, että 2. tavun vokaali on todennäköisesti ollut nykyistä väljempi myös sananloppuisessa asemassa ja että *kivi*- ja *laki* : *laen* -tyyppisten sanojen nominatiivi on ollut muotoa **kive, *lake* (merkillä * eli asteriskilla kuvataan rekonstruoitua muotoa, jota ei ole elävässä kielessä tavattu). Tällöin koko sanan taivutusparadigma on ollut läpinäkyvän agglutinatiivinen eli koostunut helposti toisistaan erottuvasta sanavartalosta ja affikseista, joista kullakin on ollut oma merkityksensä. Sanan taivutus on perustunut yksinomaan sijasuffiksin liittämiseen sanavartaloon (**kive* : *kive-n*, **lake* : *lake-n*). Voidaankin olettaa, että vokaali on supistunut jäätyään painottomaan tavunloppuiseen asemaan ja että näin on syntynyt nykyinen taivutus *kivi* : *kiven*. Tästä voidaan edelleen johtaa oletus, että vaihtelemattoman *i:n* sisältävät sanat ovat *i~e*-päätteisiä sanoja nuorempia, koska niissä esiintyy suppeampi, alkuaan vain nominatiivimuodoissa syntynyt vokaali kaikissa taivutusmuodoissa. Taivutustyypin nuoruus voidaan havaita myös siitä,

että siihen kuuluvat sanat ovat usein uusia ruotsalaisperäisiä lainoja, vrt. *lasi* < *glas*, *laki* < *lag* jne. Nykyisin useimmat kieleen saatavat lainat omaksutaan tähän tyyppiin, vrt. *traktori*, *professori*, *raportti*, *konteksti*.

Sanoissa tyyppiä *käsi* ja *tosi* esiintyy nominatiivimuodossa *s*-äänne, joka vaihtelee muiden taivutusmuotojen *t*- ja *d*-äänteiden kanssa. *s*-äänne esiintyy ainoastaan *i*-äänteen edellä, mikä antaa aiheen olettaa, että se on syntynyt *t*-äänteestä tässä asemassa. Kyseessä on siis *ti* > *si* -muutoksen myötä syntynyt *si*-sekvenssi. Se, että asia todella on näin, voidaan havaita vertaamalla suomen *si*-sekvenssin sisältämiä sanoja samanmerkityksisiin sukukielten sanoihin, vrt. *siivota* ~ pohjoissaame *divvut* 'järjestää', *sika* ~ ersämordva *tuvo*.

Sanoissa *käsi* ja *tosi* havaitaan edelleen myös *t*- ja *d*-äänteen vaihtelua. Mikäli käytettävissämme olisi koko suomen murteiden aineisto, havaitsisimme, että suomen kirjakielen *d*-äänteen asemassa esiintyvät murteissa mm. äänteet *r*, *l* ja kato (*käsi* : *kären*, *käsi* : *kälen*, *käsi* : *käen*). Vanhastaan on samassa asemassa esiintynyt lähes kadonnut dentaalispirantti *δ*, joka oli vielä Agricolan aikana käytössä myös kirjoitetussa kielessä. Vertaamalla näiden äänteiden esiintymistä *t*-äänteeseen voimme havaita, että ne esiintyvät erilaisessa fonologisessa kontekstissa: *t* esiintyy tavuissa, jotka muodostuvat konsonantista ja sitä seuraavasta vokaalista (ns. avotavuissa), kun taas *d* ja muut sen asemassa esiintyvät äänteet esiintyvät konsonanttiin päättyvissä tavuissa (ns. umpitavuissa). Todellisuudessa esiintymistä säätelevät säännöt ovat vielä hieman monimutkaisempia.

Kyseessä on suomen kielen *astevaihteluna* tunnettu avo- ja umpitavujen edellä esiintyvää konsonantismia koskeva vaihtelu. Se koskee suomen kielessä kaikkia klusiileja (*k*, *t*, *p*) sekä niiden muodostamia geminaattoja (*kk*, *tt*, *pp*) ja monia klusiileja sisältäviä äänneyhtymiä (*rk*, *lk*, *nk* jne.). Astevaihtelua esiintyy suomen kielessä sekä sanavartaloissa (*takka* : *takan*) että suffikseissa (*kuusikko* : *kuusikon*).

Myös astevaihtelun tapauksessa voidaan turvallisesti olettaa, että umpitavun etinen "heikko" ääntämys (*d*, *r*, *l* jne.) edustaa sekundaaria kehitystä, jonka taustalla on pyrkimys helpompaan ääntämiseen, ja että avotavun aloittava *t* on näissä sanoissa alkuperäinen. Tässäkin tapauksessa asia voidaan varmistaa tutustumalla suomen sanojen vastineisiin

sukukielissä: *tosi* ~ pohjoissaame *duohta, käsi* ~ pohjoissaame *giehta,* ersämordva *ked'*).

Kun tunnemme äännevaihtelujen ehdot, voimme rekonstruoida suomen kielen edeltäjään seuraavan vaihteluttoman järjestelmän:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *lasi:* | *lasin:* | *lasia* | *kive:* | *kiven:* | *kiveä* | *käte:* | *käten:* | *käteä* |
| *laki:* | *lakin:* | *lakia* | *lake:* | *laken:* | *lakea* | *tote:* | *toten:* | *totea* |

Tästä järjestelmästä voidaan selittää nykysuomen taivutus muutamilla säännöllisillä äännelaeilla, jotka ovat seuraavat:

*e#* > *i#* (sanan lopussa)

*ti* > *si*

*t, k, p* > *d, ø, v* _VC (klusiili heikkenee konsonanttiloppuisen tavun alussa).

Edellä esitetty järjestelmä ei ole täysin sama kuin se, joka todellisuudessa rekonstruoidaan suomen kielen historialliseen edeltäjään kantasuomeen. Siitä puuttuu eräitä morfologisia rajoituksia, kuten se, että possessiivisuffiksien kanssa suomen kielessä esiintyy ainoastaan ns. vahva aste (*käteni, kätesi, kätemme, kätensä* jne.). Kielessämme on myös ns. suffiksaalinen astevaihtelu eli 2. ja 3. tavun välissä havaittava astevaihtelu, joka liittyy mm. partitiivimuotoihin (vrt. *lasia, lohta, jumalaa, kiveä, luuta*).

Todellisuudessa kantasuomen järjestelmä on ollut seuraavanlainen:

| | | | | | |
|---|---|---|---|---|---|
| *kive:* | *kiven:* | *kivetä* | *käte:* | *käten:* | *kätetä* |
| *lake:* | *laken:* | *laketa* | *tote:* | *toten:* | *toteta* |

(sanat *lasi* ja *laki* ovat kantasuomen järjestelmää nuorempia).

Sisäistä rekonstruktiota voidaan käyttää suomen lisäksi minkä tahansa morfologialtaan rikkaan kielen tutkimiseen. Esimeriksi venäjän kielessä esiintyy äännevaihteluja saman sanan eri taivutusmuodoissa. Osaamista ja voimista merkitsevän verbin infinitiivimuoto on *moč*, yksikön 1. persoonan muoto *magú* ja yksikön 2. persoonan muoto *móžeš* (venäjän ortografian mukaan мочь : могу : можешь). Puhtaasti synkronisesti voidaan havaita, että vokaalivaihtelu liittyy sanapainoon siten, että sama äännesegmentti, joka esiintyy painollisena *o*:na, ääntyy painottomana *a*:na. Voidaan olettaa, että näissä tapauksissa painollinen äänne edustaa alkuperäistä kantaa. *č* ja *ž* taas esiintyvät lähinnä etuvokaalin edellä, kun taas *g* esiintyy takavokaalin edellä. Takavokaalin etiset ääntämykset ovat alkuperäisiä ja ääntöpaikasta selittyvä *č* ja *ž* syntyneet sekundaaristi etuvokaalin edellä. Voidaankin olettaa, että alkuaan kaikissa taivutusmuodoissa on esiintynyt *g*-äänne, joka sitten on muuttunut *ž*- ja *č*-äänteiksi erilaisin ehdoin.

Sisäinen rekonstruktio mahdollistaa siis melko monipuolisen kuvan saamisen kielen äännehistoriasta, vaikka käytettävissä ei olisi kuin yhden kielen aineistoa. Usein sen avulla on mahdollista paitsi kuvata tapahtuneet äänteenmuutokset myös rekonstruoida niiden järjestys. Sisäistä rekonstruktiota on mahdollista soveltaa myös niihin kieliin, joilla ei ole lainkaan tunnettuja sukukieliä.

## 4.3. Ulkoinen rekonstruktio

Historiallis-vertaileva metodi ulottaa vertailunsa yksittäisten kielten muotoja laajemmalle eri kielten välille. Ulkoista rekonstruktiota voidaan havainnoida tutkimalla esimerkiksi seuraavia suomen ja viron välisiä muutoksia:

| suomi | viro | suomi | viro | suomi | viro |
|---|---|---|---|---|---|
| *kala* | *kala* | *kana* | *kana* | *sala-* | *sala-* |
| *muna* | *muna* | *sula* | *sula* | *vesa* | *võsa* |
| *orava* | *orav* | *jumala* | *jumal* | *pesä* | *pesa* |
| *veri* | *veri* | *käsi* | *käsi* | *tosi* | *tõsi* |
| *musta* | *must* | *hirvi* | *hirv* | *perä* | *pera* |
| *ovi* | *õu* | *pensas* | *põõsas* | *kansi* | *kaas* |

Voidaan havaita, että sanat vastaavat toisiaan eri tavoin riippuen niiden äännerakenteesta. Sanat, joissa 1. ja 2. tavun välillä on yksittäiskonsonantti, ovat suomessa ja virossa lähes identtisiä. Niissä sanoissa taas, joissa 1. ja 2. tavun välillä on ollut konsonanttiyhtymä, vastaavat suomen kaksitavuisia sanoja viron kielessä yksitavuiset sanat. Aineiston pohjalta voimme päätellä, että viron kielessä on tapahtunut sananloppuinen muutos, vokaalin katoaminen, joka riippuu edeltävän konsonantismin luonteesta. Vastaava muutos on tapahtunut myös 3. tavussa yksittäiskonsonantinkin jäljessä.

Aineistosta näkee myös monia muita muutoksia. Esimerkiksi suomen 2. tavun *ä*-äännettä vastaa viron 2. tavun *a*-äänne. Mikäli taas suomessa on ollut 1. tavussa *e* ja 2. tavussa *a* kuten sanoissa *vesa* ja *pensas,* on virossa syntynyt 1. tavuun äänne *õ* (keskivokaali, joka äännetään samalla korkeudella kuin suomen takavokaali *o* mutta huulet leveinä kuten *e*:tä äännettäessä).

Jos oletetaan, että suomen kanta on alkuperäinen ja viron uudempi, voidaan edellä olevan aineiston pohjalta kirjoittaa seuraavat virossa toteutuneet äännelait:

1) (C)VCCV# → (C)VCCØ# (kaksitavuisissa sanoissa on 2. tavun vokaali kadonnut kaksoiskonsonantin jälkeisessä asemassa)
2) (C)VC(C)*ä* → (C)VC(C)*a* (sananloppuinen -*ä* on muuttunut *a*:ksi)
3) (C)*e*C(C)*a* → (C)*õ*C(C)*a* (jos 1. tavussa on ollut *e* ja 2. tavussa *a*, on 1. tavun *e* muuttunut õ-äänteeksi)

(C = konsonantti, V = vokaali, # = sananloppu, Ø = kato)

Oletus, että viron sanoissa on tapahtunut muutoksia, kun taas suomen sanat edustavat alkuperäistä kantaa, perustuu myös muihin sukukieliin. Näistä useimmat ovat säilyttäneet vokaalisoinnun (*perä*-tyypin vartalot) ja kaksitavuiset vartalot *musta*-tyypin sanoissa. Sen sijaan kysymys *õ*-äänteen alkuperästä on mutkikkaampi. Aikaisemmin on oletettu melko yksimielisesti, että suomen kieli edustaa tässäkin alkuperäistä kantaa, mutta mm. Petri Kallio (2012) on puolustanut käsitystä, jonka mukaan redusoitunut vokaali on kuulunut kantasuomen äännejärjestelmään ja kadonnut suomessa. Suomen *vesa* edustaisi siis uutta kehitystä ja viron *võsa* alkuperäista sanamuotoa.

## 4.4. Äännekontekstien vastaavuus

Todellisuudessa historiallis-vertailevaa metodia kuitenkin sovelletaan vertailemalla lukuisia kieliä yhtä aikaa, jolloin yhtäläisyydet alkavat näyttää ymmärrettävämmiltä. Aineisto voi näyttää esimerkiksi tältä:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| suomi | *p* | *ä* | | *ä* | *y* | *d* | *i* | *n* | |
| saame | *b* | *á* | *kŋ* | *i* | *a* | *đ* | *a* | | |
| ersä | *p* | *e* | | | *u* | *d'* | *e* | *m* | *e* |
| mansi | *p* | *ä* | *ŋk* | | *v* | *ä* | *l* | *ø* | *m* |
| unkari | *f* | *e* | *j* | | *v* | *e* | *l* | *ő* | |
| komi | *p* | *o* | *m* | | *v* | *e* | | | *m* |
| suomi | *p* | *i* | | *i* | *k* | *ä* | *s* | *i* | |
| saame | *b* | *a* | *tn* | *i* | *g* | *ie* | *ht* | *a* | |
| ersä | *p* | *e* | *j* | | *k* | *e* | *d'* | | |
| mari | *p* | *ü* | *j* | | *k* | *i* | *d* | | |
| komi | *p* | *i* | *n'* | | *k* | *i* | | | |
| hanti | *p* | *e* | *ŋk* | | *k* | *a* | *t* | | |
| unkari | *f* | *o* | *g* | | *k* | *é* | *z* | | |

Esimerkin asettelusta näkyy, että historiallis-vertaileva metodi ei tarkkaan ottaen vertaile äänteitä eikä sanoja vaan sanoissa esiintyviä toisiaan vastaavia äännesekvenssejä (vrt. erit. Fox 1995). Monissa tapauksissa toisiaan saattavat vastata esimerkiksi kato ja yksittäisäänne tai useamman äänteen kokonaisuus. Juuri monimutkaisten konteksti kontekstilta toistuvien äännevastaavuuksien tähden voimme olla varmoja siitä, että kielten historiallinen kehitys ei ole sattumanvaraista.

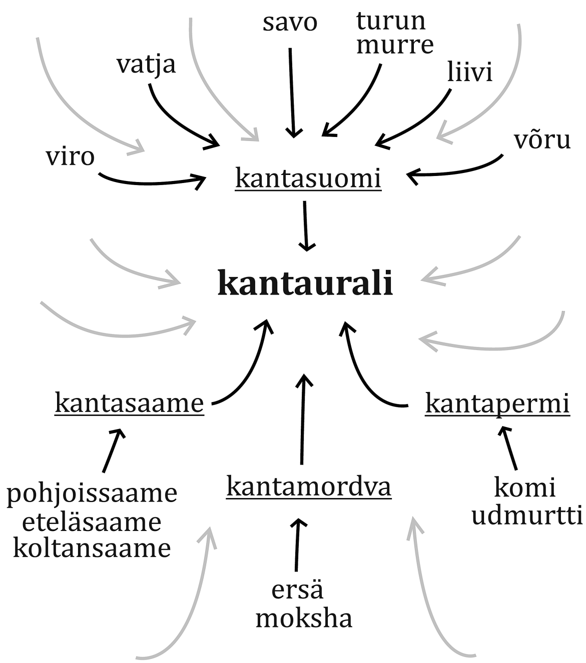
Esimerkiksi suomen sanassa *pii* ja unkarin sanassa *fog* 'hammas', tai suomen sanassa *ydin* ja unkarin samaa tarkoittavassa sanassa *velő* ei ole yhtään samaa äännettä. Kuitenkin kun vertaamme kaikkien sukukielten aineistoa, sanojen yhteenkuuluvuus alkaa näyttää ilmeiseltä. Uralilaisissa kielissä on monia sanoja, joissa alkuaan hammasta tarkoittanut *pii*-sana on konsonanttipäätteinen tai kaksitavuinen. Tämä osoittaa, että suomen sanassa on tapahtunut keskikonsonantiston kato, joka on johtanut sanan muuttumiseen yksitavuiseksi. Historiallis-vertailevan metodin avulla voidaan helposti todeta, että vastaava muutos on tapahtunut muissakin yksitavuisissa pitkän vokaalin sisältävissä sanoissa, siis myös esim. sanoissa *luu*, *jää*, *pyy*, *sää* ja niin edelleen.

Äänteiden katoaminen selittää hyvin, miksi suomen *ydin* ja unkarin *velő* ovat sama sana. Vaikka sanat näyttävät näennäisesti aivan erilaisilta, ne noudattavat aivan samoja äännevastaavuuksia kuin monet muut suomen *y*-alkuiset sanat, *n*-päätteiset sanat tai sanat, joissa esiintyy *d* 2. tavun alussa. Usein amatöörimäiset kielivertailut tunnistaakin siitä, että ne keskittyvät samannäköisiin sanoihin sen sijaan, että yritettäisiin osoittaa sanojen välillä vallitsevia systemaattisia äänteellisiä yhteyksiä. Esimerkiksi melko paljon julkisuutta saanut suomen ja sumerin sanaston vertailu (Parpola 2016) edustaa tällaista etäsukulaisuushypoteesia, joka ei toteuta normaalin kielisukulaisuuden perusvaatimusta eli äänteellistä säännönmukaisuutta (Saarikivi 2016).

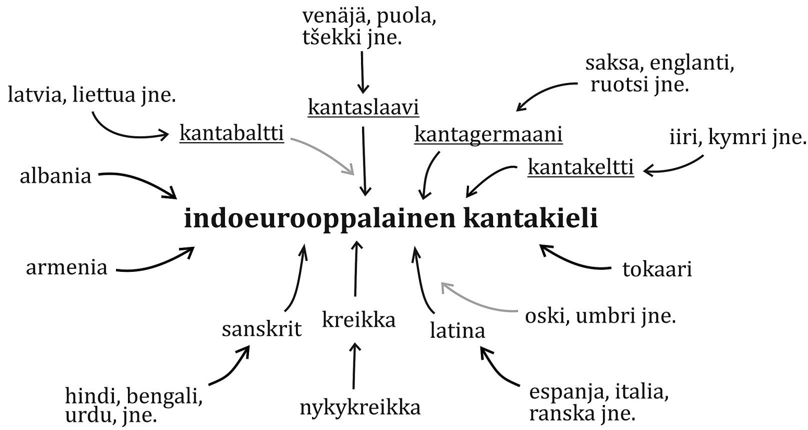
## 4.5. Historiallis-vertaileva metodi ja kielten taksonomia

Historiallis-vertailevaan menetelmään kuuluu oleellisesti kielikunnan taksonomia eli kysymys siitä, mikä kieli polveutuu mistäkin

rekonstruktiotasosta ja millä tavoin nykykielet suhteutuvat toisiinsa. Esimerkiksi suomi ja viro palautuvat kantasuomeen, joka puolestaan palautuu kantauraliin. Vastaavasti venäjä ja puola palautuvat kantaslaaviin, joka palautuu indoeurooppalaiseen kantakieleen. Suomalais-ugrilainen kielikunta ja indoeurooppalainen kielikunta ovat molemmat kaksitasoisia siinä mielessä, että niissä on yhtäältä sekä nykyisten tai kirjallisesti tunnettujen kielten pohjalta rekonstruoituja kantakieliä (kuten kantasuomi, kantasaame, kantagermaani, kantaslaavi jne.) ja toisaalta näitä kantakieliä vertailemalla luotu koko kielikunnan yhteinen kantakieli. Tällaista monitasoista rekonstruktioprosessia voidaan kuvata seuraavanlaisilla kaavioilla:

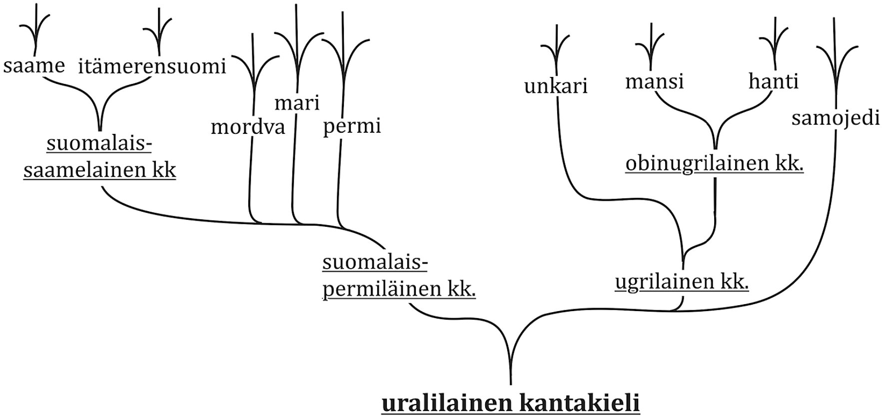



Kuvio 1. Kantauralin rekonstruktioprosessi (kuvion Copyright Nora Fabritius). Koko kielikunnan kantakielen rekonstruktio luodaan uudempien kantakielten rekonstruktioita vertailemalla.

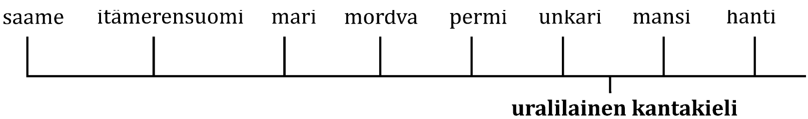



Kuvio 2. Indoeurooppalaisen kantakielen kaksivaiheinen rekonstruktioprosessi (kuvion Copyright Nora Fabritius).

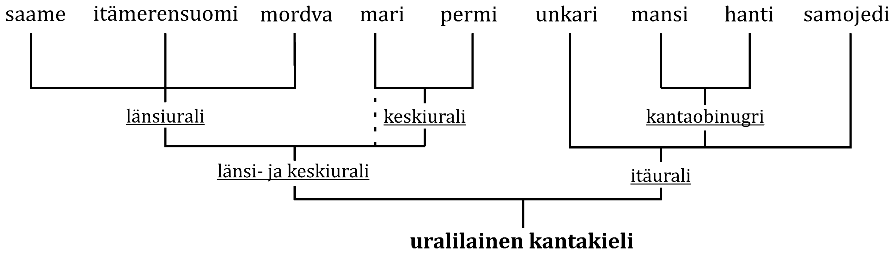
Kantaurali (kuvio 1) tai indoeurooppalainen kantakieli (kuvio 2) ovat siis toisen asteen rekonstruktioita, jotka on luotu rekonstruktioita vertailemalla. Kielikuntien **sukupuut** kuvaavat sitä, missä suhteessa eri-ikäiset kantakielet ovat toisiinsa ja minkä tason rekonstruktiot ovat kyseessä. Esimerkiksi suomalais-ugrilaisten kielten tapauksessa on keskusteltu ns. binaarisesti jakaantuvan ja pensasmaisen sukupuun eroista (kuvio 3). Binaarinen sukupuu väittää, että suomen kielen edeltäjiä ovat kantasuomi, varhaiskantasuomi (itämerensuomalaisten kielten ja saamen yhteinen kantakieli), suomalais-permiläinen kantakieli (itämerensuomalaisten kielten, saamen, mordvan, marin ja permiläisten kielten kantakieli), suomalais-ugrilainen kantakieli (itämerensuomalaisten kielten, saamen, mordvan, marin ja permiläisten kielten kantakieli ja ugrilainen kantakieli) ja kantaurali tässä järjestyksessä. Pensasmallinen sukupuu (kuvio 4) taas esittää, että suomen kielen kantamuoto kantasuomi on vain yksi kantauralin lukuisista tytärkielistä. Useimmat tutkijat ovat taipuvaisia ajattelemaan, että niin suomalais-ugrilaisten kuin indoeurooppalaisten kielten sukupuu muistuttaa pikemminkin pensasta kuin puuta, mutta monet yksityiskohdat ovat hämäriä (kuviot 5–7).



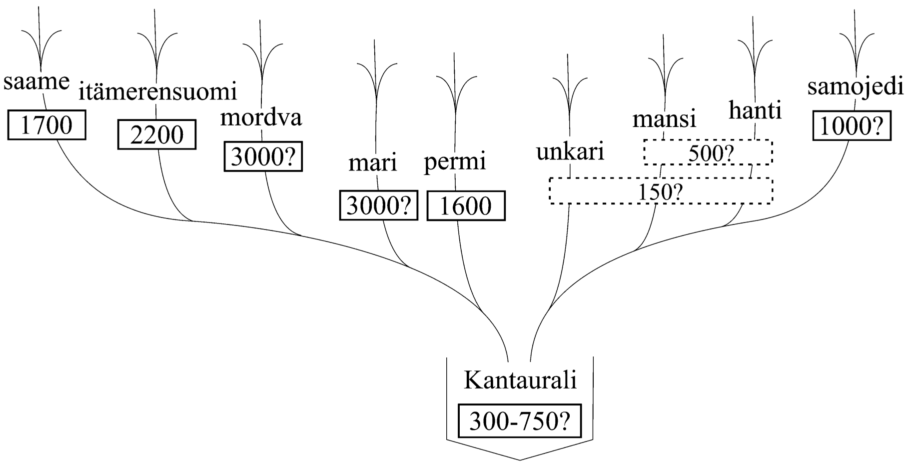

Kuvio 3. Binaarisesti haarautuva uralilaisten kielten sukupuu Kulosen (2002) mukaan (kuvion Copyright Nora Fabritius).



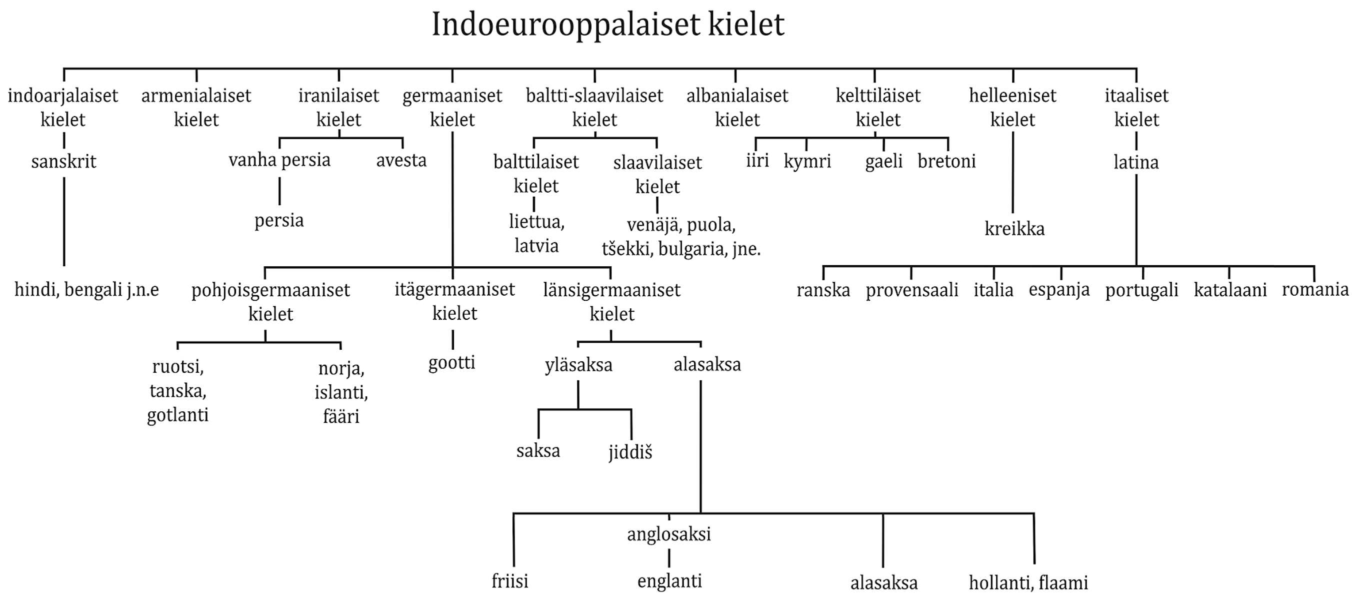

Kuvio 4. Tapani Salmisen (1999) mukainen sukupensas (kuvion Copyright Nora Fabritius). Tällainen sukupuumalli olettaa kantakielen jakaantuneen useisiin kantakieliin jo varhaisvaiheessaan.




Kuvio 5. Häkkisen (2007) mukainen ehdotus sukupuuksi (kuvion Copyright Nora Fabritius).




Kuvio 6. Uusin uralilaisten kielten sukupuu (Saarikivi, painossa). Sulkeissa kuhunkin kantakieleen palautuvan yhteissanaston määrä (kuvion Copyright Nora Fabritius).




Kuvio 7. Indoeurooppalaisten kielten sukupuu (kuvion Copyright Nora Fabritius).

# 5. Etymologia

## 5.1. Sanojen alkuperän päätyypit

Edellä on käsitelty äänteiden ja sanojen säännönmukaista kehitystä. Kielten sanasto voi olla paitsi kantakielestä perittyä myös johdettua, yhdistettyä tai toisista kielistä lainattua. Usein erotetaan vielä äänteellisesti motivoitu sanasto, ns. deskriptiiviset ja onomatopoieettiset sanat (vrt. esim. Häkkinen 1990) eli äänneasullaan asioita kuvaavat tai ääntä jäljittelevät sanat.

Kun sanoja tutkitaan etymologisesti, käytetään peruskriteereinä yleensä sanan äänneasua, jonka tulisi systemaattisesti vastata joko sukukielten tai lainantajakielen äänneasua, sanan merkitystä, jonka tulisi olla sama tai läheinen kuin sukukielten tai lainanantajakielen sanan (ks. tarkemmin alla), ja sanan levikkiä, jonka tulisi viitata joko laina-alkuperään (mikäli sana esiintyy naapurikielissä) tai periytymiseen kantakielestä (mikäli sana esiintyy sukulaiskielissä) (ks. tarkemmin Häkkinen 1990).

## 5.2. Perintösanasto

Edellä esitetyissä esimerkeissä on ollut puhe pääasiassa kantakielestä periytyneestä sanastosta. Tällaisella sanastolla ei tavallisesti ole muuta tunnettua alkuperää. Perintösanastoon eli kielen kauimmin säilyvään sanastoon kuuluu tyypillisesti ruumiinosien nimityksiä (*käsi*, *pää*, *veri*, *maksa*), sääsanastoa (*tuuli*-, *kylmä*, *lumi*), luonnon perusilmiöiden nimityksiä (*vesi*, *puu*, *maa*, *kivi*) ja ihmisen perustoimintoja (*mene*-, *näke*-, *kuule*-, *tule*-) kuvaavia sanoja. Teknologioihin, raaka-aineisiin, uskontoon ja kulttuuriin tai paikallisesti rajattuihin ilmiöihin liittyvä sanasto taas on todennäköisemmin lainasanastoa (ks. luku 5.3.).

Historiallis-vertailevaa metodia sovellettaessa oletetaan, että äännelaillisesti toisiaan vastaavat sanat ovat olleet olemassa kantakielessä ja että niitä on käytetty jatkuvasti siten, että ne ovat muuttuneet nykykielten sanoiksi. Historiallisessa dimensiossa esimerkiksi suomen *ydin* ja unkarin *velő* on siis yksi ja sama sana.

## 5.3. Lainasanat

Yleensä kaikissa kielissä on ainakin jonkin verran toisista kielistä lainattua sanastoa, vaikkakin lainasanaston määrä vaihtelee hyvin paljon. Esimerkiksi kiinan kielen sanastosta alle 1 % on voitu osoittaa lainasanoiksi (Haspelmath & Tadmor 2009), kun taas vaikkapa suomen tai englannin perussanoista valtaosa on eri-ikäisiä lainoja.

Eri aikoina on suomen kielen edeltäjiin päätynyt mm. maanviljelyyn liittyvää sanastoa germaanisista kielistä (_pelto_ < germaaninen _*feldō_ [vrt. englannin _field_], _aura_ < germaaninen _*arðra_ [vrt. ruotsin _årder_ ’aura’]) ja metallien nimityksiä (_rauta_ < germaaninen _*rauða_ > ruotsin _röd_ ’punainen’, alkuaan myös ’malmi’, _hopea_ < _*šompeδa_ < germaaninen _*somptjo_ ’pehmeä’ > ruotsi, englanti _soft_) sekä kristinuskoon liittyvää sanastoa slaavilaisista kielistä (_raamattu_ < kantaslaavin _grāmatā_ ’teksti; kirjoitus’ > nykyvenäjän _gramota_ ’kunniakirja’, _pappi_ < kantaslaavin _papŭ_ > nykyvenäjän _pop_, _risti_ < kantaslaavin _kristŭ_ > nykyvenäjän _krest_).

Lainasanoja otetaan usein sellaisiinkin merkityskenttiin, joissa on jo vanhaa sanastoa. Esimerkiksi suomessa on monia myöntämiseen käytettyjä sanoja (_joo_, _jaa_, _kyllä_), mutta tämä ei ole estänyt sellaisten lainasanojen kuin _okei_ ja _jes_ leviämistä moderniin puhesuomeen. Vastaavanlaista muoti-ilmiöihin liittyvää lainaamista on selvästi ollut jo varhaisina aikoina. Esimerkiksi balttilaisista kielistä lainattiin suomeen mahdollisesti pronssikaudella sanat _hammas_ ja _kaula_, jotka korvasivat sanat _pii_ (esim. ilmauksissa _haravan pii_ ’haravan hammas’ tai _piikivi_ ’terävä hammasreunainen kivi’) ja _sepä_ (esim. johdoksessa _sepalus_).

Kielestä toiseen lainattaessa lainansaajakielestä puuttuvat äänteet ja äänneyhtymät mukautetaan lainansaajakielen äännejärjestelmään ja fonotaksiin, eli ne käyttäytyvät lainansaajakielen äänteiden yhdistämissääntöjen mukaisesti. Tällaista mukauttamista kutsutaan _äännesubstituutioksi_. Esimerkiksi ruotsin kielen sananalkuiset konsonanttiyhtymät on yleensä korvattu suomen kielessä yksittäiskonsonantilla, kuten tapauksissa _laji_ < _slag_, _talli_ < _stall_, _lankku_ < _planka_, _lantti_ < _slant_, _raapia_ < _skrapa_, _rahti_ < _frakt_, _Ranska_ < _fransk(a)_ ym. Vastaavasti esimerkiksi unkarin kielessä on sananalkuiset konsonanttiyhtymät usein mukautettu kieleen lisäämällä uusi vokaali sanan alkuun: latinan _schola_

> *iskola* ’koulu’, italian *stallo* > *istalló* ’talli’, saksan *Spital* > *ispotaly* (vanh.) ’sairaala’.

Tutkimus on osoittanut, että samalla tavoin kuin kielen sisäiset äännekehitykset myös lainaaminen on säännönmukaista. Samassa lainautumistilanteessa on yleensä käytössä samantyyppiset lainasubstituutiot kaikkien sanojen ja kaikkien samanlaisten äännekontekstien tapauksessa. Sama äänne voidaan silti substituoida hyvinkin eri tavoin riippuen siitä, esiintyykö se painollisena vai ei, minkä äänteen vieressä se esiintyy, tai siitä, onko kyseessä sanan loppu- tai alkuäänne jne. Paljon nykykielissä ilmenevästä lainasanaston äänteellisestä moninaisuudesta johtuu myös siitä, että sanat on alkuaan lainattu eri murteisiin ja eri aikoina eri lähteistä.

Esimerkiksi sananalkuinen *st*-yhtymä näyttää pitkään lainautuneen suomen kielen edeltäjiin *s*-äänteenä. Tämä näkyy lainasuhteissa indoiraaninen **stambhas* > *sammas* ’pylväs’, kantabaltti **staibis* > *seiväs*, kantagermaani **stūra-* > *suuri* (SSA [nide 3], 125, 139, 200; SKES: 962, 992, 1136–1137). Jossain vaiheessa substituutiomalli tai -muoti on kuitenkin muuttunut ja *st*-yhtymää on alettu korvata *t*-äänteellä (vrt. edellä). Sanan ikää voidaan näin ollen arvioida sen mukaan, millaisia lainasubstituutioita siinä on käytetty.

Sanan ikää voidaan arvioida myös sen perusteella, millainen on ollut lainasanan osoittama lainaoriginaali, ts. se lainanantajakielen sanahahmo, josta nykykielen sana on saatu. Esimerkiksi germaanisten kielten ’kuningasta’ merkitsevät sanat ovat nykyisin ruotsin *kung*, saksan *König*, englannin *king* jne. Varhaisemmissa kirjoitetuissa kielimuodoissa sana on kuulunut *konungr* (muinaisnorja), *cuning* (anglosaksi) ym. Kaikki nämä muodot voidaan johtaa kantagermaanin rekonstruoidusta muodosta **kuningaz*, joka taas näyttää olevan suomen sanan *kuningas* alkumuoto. Äänteellisesti läheinen on myös liettuan pappia merkitsevä *kunygas*. Nämä kaksi sanaa on omaksuttu sellaisesta germaanin muodosta, joka on kirjallisesti säilyneitä kielimuotoja äänteellisesti vanhakantaisempaa. Ne todistavat osaltaan sanan historiallis-vertailevaan metodiin perustuvan rekonstruktion olevan uskottava.

Äänteellisen mukautumisen lisäksi lainasanasto tulee mukauttaa lainansaajakieleen myös morfologisesti. Jos esimerkiksi lainansaaja-

kielessä on kieliopillisia sukutunnuksia, saavat myös lainasanat yleensä sellaiset. Vastaavasti suomen kieleen omaksutut sanat noudattavat jotakin suomen kielen monista taivutustyypeistä. Melko usein lainasanat saavat samanlaisia suffiksaalisia päätteitä kuin vanha peritty sanasto.

Suomen kielellä on lainasanatutkimuksen historiassa keskeinen rooli, sillä se näyttää säilyttäneen monia indoeurooppalaisten kielten varhaisia sanahahmoja hyvin. Suomen kielen muutosnopeus näyttää olleen suhteellisen hidas, ja kieli on siinä mielessä äänteellisesti konservatiivista. Ensimmäiset merkittävät lainasanatutkimukset germaanisten ja slaavilaisten kielten alalla, kuten Vilhelm Thomsenin (1870) ja Jooseppi J. Mikkolan (1894) tutkimukset, käsittelivät juuri lainakontakteja suomen kielen esimuotojen kanssa.

1900-luvun jälkipuoliskolla erityisesti Jorma Koivulehto tutki suomen kontakteja varhaisiin germaanisiin, balttilaisiin ja muihin indoeurooppalaisiin kielimuotoihin. Hänen tutkimuksensa vaikuttivat monin tavoin kuvaan Euraasian kielellisestä esihistoriasta ja myös indoeurooppalaisten kielten historiasta. Koivulehto on jopa ehdottanut, että suomen kielessä olisi lainasanastoa, joka todistaisi kantaindoeuroopan ns. laryngaaliäänteistä eli kurkkuäänteistä, jotka sellaisenaan eivät ole säilyneet missään indoeurooppalaisessa kielessä heettiä lukuun ottamatta (Koivulehto 1991), mutta tätä hypoteesia ei ole yksimielisesti hyväksytty.

## 5.4. Äänteellisesti motivoidut sanat

Äänteellisesti motivoitunutta sanastoa lienee kaikissa kielissä. Se ei kuitenkaan ole niissä samanlaista, sillä se ei jäljittele ääniä sellaisenaan vaan kunkin kielen fonemaattisiin konventioihin liittyen (Aikio 2001; Saarikivi 2006). Esimerkiksi suomen kielessä on olemassa konventio, jonka mukaan johtimellinen verbi, jossa on keskikonsonantistossa -*r*- tai -*l*-, voi olla ääntä jäljittelevä, kuten sanat *surista*, *hurista*, *korista* ja *piristä*. Tämän äänteellisen konvention mukaisesti on tarvittaessa mahdollista luoda uusia ääntä jäljitteleviä sanoja. Toisaalta esimerkiksi sellaiset sanat kuin *varis*, *kaarne* tai *korppi* vaikuttavat alkuaan ääntä jäljitteleviltä, mutta niiden taustalla oleva äänteellinen konventio näyttää kadonneen

kielestä. Nykykielen kannalta kyseessä ovat siis neutraalit sanat. Helposti ei huomaa sitäkään, että esimerkiksi sanat *kapina*, *huhu* ja *juoru* ovat aikoinaan olleet äänteellisesti motivoituja.

Ääntä jäljittelevä sana voi edelleen muuttua deskriptiiviseksi. Tämä tarkoittaa, että sen ääntä jäljittelevä merkitys jää taka-alalle. Esimerkiksi sellainen sana kuin *pimahtaa* on alkuaan syntynyt ääntä jäljittelevästä tavusta *pim*. Nykykielessä se tarkoittaa kuitenkin lähinnä suuttumista tai järjen menettämistä. Kyseessä on alkuaan äänen metaforinen tulkinta: ihminen käyttäytyy ikään kuin hänen päässään kuuluisi ”pim”.

## 5.5. Substraatti

*Substraattivaikutuksella* tarkoitetaan kuolleiden tai alueellisesti väistyneiden kielten vaikutusta nykykieliin, esim. saamelaista vaikutusta Peräpohjolan murteissa ja laajemminkin suomessa (Aikio 2008) tai itämerensuomalaista vaikutusta pohjoisvenäläisissä murteissa (Saarikivi 2006), sekä myös kielenvaihdon yhteydessä tapahtuvia puutteellisen kielenoppimisen vaikutuksia (ks. Kielikontaktien tutkimus tk.). Nämä sanan *substraatti* kaksi merkitystä ovat sillä tavoin erilaisia, että edellinen kuuluu varsinaiseen historiallis-vertailevaan kielentutkimukseen, jälkimmäinen kontaktilingvistiikkaan.

Historiallisesti säilyvät substraattivaikutukset ovat tyypillisesti esim. paikannimistöä tai sanastoa (Saarikivi 2006). Usein myös kieliopissa ja kielen fonetiikassa on havaittavissa substraattivaikutuksia, joiden todistaminen voi kuitenkin olla vaikeaa. Tämä johtuu siitä, että kielen muutoksen mekanismeja on monia erilaisia ja ne ovat kontekstista toiseen vaihtelevia (ks. Kielikontaktien tutkimus tk.).

Substraattisanasto on lainasanastoa siinä mielessä, että se on saatu toisista kielistä, mutta kahden erikielisen ryhmän kontaktien asemesta sanaston omaksumisprosessi on ollut pikemminkin kielenvaihto ja kielikuolema sekä tiettyjen sanojen säilyminen sen yhteydessä. Esimerkiksi saamelaiskielissä on runsaasti sanastoa, joka liittyy pohjoiseen luontoon ja elinkeinoihin mutta joka ei periydy varhaisemmista kantakielistä tai ole tunnettua lainaa naapurikielistäkään, esim. pohjoissaamen *njálla*

’naali’, *morsa* ’mursu’, *gáisa* ’lumihuippuinen vuori’, *jássa* ’lumipälvi vuoren rinteellä’ jne. Tällöin on ilmeinen johtopäätös, että sanasto on saatu alueella aiemmin puhutuista kielistä substraattivaikutuksina.

### 5.6. Kontaminaatio ja kansanetymologia

Sanaston historiaan liittyy usein monenkertaisia johtoprosesseja sekä sanojen sekoittumisia eli kontaminaatioita. Esimerkiksi sana *kolaholisti* on kontaminaatio, joka on muodostettu sanojen *alkoholisti* ja *kola* pohjalta. Usein kontaminaatioita aiheuttaa kansanetymologia eli ääneasultaan läheisten sanojen samaistaminen toisiinsa. Kansanetymologia on analogian erityinen muoto, jota havainnollistaa Helsingin *Huopalahden* kaupunginosan nimi. Alun perin alueen nimi on ollut *Haapalaksi,* joka on lainautunut ruotsiin asussa *Hapalax*. Tämä on ruotsin kielessä tapahtuneiden muutosten ansiosta muuttunut nykyasuunsa *Hoplaks,* joka sitten on jälleen lainattu suomeen muodossa *Huopalahti*. Nimellä ei ole mitään tekemistä huovan kanssa, eikä sen toinen komponentti ole alkuaan *lahti* vaan *laksi*, mutta nimi on kansanetymologisesti samaistettu fonemaattisesti läheisiin sanoihin.

## 6. Muotohistoriaa

Periaatteessa aivan samalla tavoin kuin äänteistön ja sanaston historiaa tutkitaan historiallis-vertailevan menetelmän avulla myös kielen taivutusmuotojen eli morfologian historiaa. Esimerkiksi jo edellä on viitattu astevaihteluun eli suomen kielen klusiilien heikkenemiseen historiallisen umpitavun edellä. Vaikka tämä muutos on lähtökohdiltaan fonemaattinen ja koskee siis tiettyjen äänteiden muuttumista tietyssä ympäristössä, on se käytännössä muuttanut suomen kielen taivutusjärjestelmää siten, että vaikkapa äänteellisesti läheisten sanojen *pata* ja *data* genetiivimuodot muodostetaan nyt eri tavoin.

Edellä on mainittu kieliopillistumisesta eli grammatikalisaatiosta. Kielen historiassa toistuu tilanne, jossa itsenäiset sanat muodostuvat kieliopillisiksi affikseiksi, tai tilanne, jossa affiksin merkitys muuttuu. Esimerkiksi suomen kielen sanamuodosta *kansassa* on syntynyt postpositio *kanssa* (*Juhan ja Joukon kanssa*, alkuaan *kansassa* eli 'Juhan ja Joukon seurassa'). Tämä on edelleen kehittynyt viron kielessä ja monissa suomen puhekielen muodoissa sijamuodoksi, jota voi kutsua *komitatiiviksi* tai eräissä tapauksissa *instrumentaaliksi* (vrt. viron *sõita autoga* 'ajaa autolla', kirjaimellisesti 'soittaa auton kanssa', tai suomen *juon kahvia maidonkaa*, jossa -*kaa* muistuttaa postposition asemesta sijamuotoa [Sirola-Belliard 2016]).

Sanan ja kieliopillista merkitystä kantavan affiksin välinen raja on liukuva. Yleensä affiksiksi muuttunut sana on fonemaattisesti muuttunut siinä määrin, että sitä ei enää tajuta sanamuodoksi. Tyypillistä on esim. sanapainon menetys ja monenlainen redusoituminen. Yleisesti on tapana ajatella, että säännölliset äänteenmuutokset eivät koske sellaisia sanoja, jotka ovat kieliopillistuneet eli grammatikalisoituneet, ts. muuttuneet itsenäisistä sanoista affikseiksi. Kielessä on edelleen sana *kansa*, joka on äänteellisesti säännöllinen suhteessa sukukielten ja lainanantajakielen sanoihin, mutta tämä säännönmukaisuus ei koske sanan *kansa* taivutusmuodosta *kansassa* muodostunutta postpositiota tai sijamuotoa -*ga*. Toisaalta vastaavaa redusoitumista näyttää tapahtuvan ainakin jonkin verran myös yhdyssanoille ja johdoksille niiden merkityksen muuttuessa. Esimerkiksi suomen *taas* on alkuaan sama sanamuoto kuin *takaisin*, ja suomen *tänään* on alkuaan sanaliitto *tänä päivänä*.

Kieliopillisten taivutusmuotojen kehityksen edellytys on sanojen ja affiksien merkityksen muutos ns. reanalyysin kautta. Yleensä sanan ja suffiksin raja ymmärretään uudelleen siten, että osa vanhasta sanavartalosta siirtyy suffiksiin. Esimerkiksi savolaismurteissa tavattava *loi*-monikko (*kukkoloita*, *taloloita*) on syntynyt ilmeisesti sanoista, joissa -*l*- on alun perin kuulunut sanavartaloon (vrt. *pappila* : *pappiloita*, *satula* : *satuloita*). *l*-aines on kuitenkin tajuttu päätteen osaksi, ja sitä on alettu yleistää kaikenlaisiin konteksteihin samaan tapaan kuin imperfektin *si*-tunnusta, josta on puhuttu edellä. Vastaavalla tavalla ovat syntyneet myös viron illatiivimuodot *koolisse* 'kouluun', *linnasse* 'kaupunkiin',

*autosse* 'autoon'. Niihin on yleistetty *se*-aineksinen illatiivin pääte niistä sanamuodoista, joissa se on säännöllinen, vrt. su *harmaa* : *harmaaseen, varis* : *varikseen,* viro *vares* : *varesse.*

# 7. Syntaksin historiaa

Historiallis-vertailevan metodin ytimessä on kielen fonologinen muoto, sanat ja affiksit. Sen sijaan se on jossain määrin heikompi kielellisten rakenteiden, kuten syntaksin, historian tutkimuksessa. Metodin tutkimuskohdetta voisi verrata esimerkiksi arkeologin esiin kaivamaan kadonneeseen taloon, jossa on historiallisten esineiden jäänteitä. Arkeologi voi tutkia esineiden tyyppejä ja muotoja. Hän voi myös pohtia, missä järjestyksessä esineet alkuaan olivat esiin kaivetussa huoneessa, mutta tämä on aivan erilainen tutkimustehtävä. Kielihistorian ytimessä taas ovat esinetyyppeihin verrattavat tutkimuskohteet: yksittäiset sanat, äänteet tai kielen muodot.

Edellä on kuitenkin jo viitattu siihen, että kielen muutokseen historiassa kuuluu myös sanojen ja affiksien funktion muuttuminen. Tämä mahdollistaa syntaksin historian tutkimuksen. Esimerkiksi suomen kielessä on demonstratiivipronominista *se* tullut eräänlainen määräinen artikkeli. Vastaava kehitys on tapahtunut myös unkarin kielessä, jossa lisäksi lukusana *egy* 'yksi' on kehittynyt epämääräiseksi artikkeliksi.

Varhaisempi samankaltainen kehitys on havaittavissa suomen verbin preesenstaivutuksessa (*minä menen, sinä menet, hän menee*), jonka 3. persoonan muodot ovat alun alkujaan olleet nominaalimuotoja. Ne ovat vastanneet nykykielen partisiippeja ja johdoksia tyyppiä *menevä, menijä*. Lauseessa *hän menee* on siis ollut ns. nominipredikaatti, ja se on tulkittu 'hän on menevä'.

Reanalyysille läheistä sukua on ilmiö, jossa affiksit saavat uusia merkityksiä, kun niitä käytetään uudessa kontekstissa. Esimerkiksi suomen genetiivin *-n*-pääte, joka on ilmaissut omistamista, on kehittynyt instruktiivin päätteeksi lauseyhteyksissä, joissa se on nominin

asemesta liittynyt verbiin (vrt. *oikean jalan kenkä → jalan kulkea*). Pääte on siis sama, mutta se saa erilaisen tulkinnan uudessa lauseopillisessa yhteydessä, ja näin syntyy uusi sijapääte. Vastaavasti monissa suomen sukukielissä, esimerkiksi permiläisissä kielissä komissa ja udmurtissa, on possessiivisuffikseista kehittynyt määräisyyden tai topikaalisuuden merkitsin, joka vastaa funktioltaan englannin määräistä artikkelia. Kun sanotaan esimerkiksi *juo tuo teesi* (komin *čaj-yd ju*!), ei possessiivisuffiksi enää viittaakaan omistamiseen vaan lauseyhteydessään määräiseen tai topikaaliseen kupilliseen teetä.

Syntaksissa kuten sanastossakin lainautuminen on yleistä, vaikka lainaperäinen rakenne tulee mukauttaa kielen järjestelmään. Esimerkiksi suomen sivulauserakenteet ovat hyvin samantyyppisiä kuin naapurikielissä ruotsissa ja venäjässä sekä balttilaisissa kielissä yhtäältä ja eroavat monista sisä-Venäjän uralilaisista sukukielistämme toisaalta. Voikin olettaa, että kyseessä ovat osittain lainaperäiset, kielikontaktissa kopioituneet rakenteet. Näihin kuuluvat sivulauseiden ilmaiseminen konjunktioiden avulla infiniittisten rakenteiden ja lauseenvastikkeiden asemesta ja ns. perifrastiset aikamuodot eli *olla*-verbistä ja verbin nominaalimuodosta koostuvat perfekti ja pluskvamperfekti.

## 8. Semanttinen muutos

Sekä periytyviin että lainautuviin kielen elementteihin liittyy paitsi äänteellinen muutos myös merkitysmuutos. Merkityksenmuutoksen tutkimusta on pitkään pidetty huomattavasti vaikeampana kuin äänteellisen muutoksen tutkimusta.

Äänteenmuutokset toistuvat samankaltaisina kielestä toiseen, mutta merkityksenmuutoksissa on paljon ainutkertaisiin kulttuurikonteksteihin liittyvää. Esimerkiksi *Casanova* oli italialainen aatelismies, mutta hänen nimestään on tullut huikentelevaa naistenkaatajamiestä tarkoittava käsite. Ruotsin *stycke* ’osa, kappale’ taas on lainattu suomeen sanana *tykki*, joka tarkoittaa ampuma-asetta. Ilmeisesti merkityksenmuutos johtuu

ranskalaisesta tavasta kutsua tykistön tykkejä 'osiksi' (*piece d'artillerie*), mikä on vaikuttanut Ruotsin armeijan puhekäytäntöihin. Ruotsin armeijassa palvelleiden tykkimiesten mukana sana on kulkeutunut suomeen alkuperäistä paljon rajatummassa merkityksessä. Vastaavia merkityksenmuutoksia, jotka voidaan ymmärtää vain kulttuurikontekstissaan, tunnetaan kaikista maailman kielistä. Eräät niistä ovat kansainvälisiä. Esimerkiksi italian sana *fiasco* 'pullo' kuvaa yleisesti epäonnistumista eurooppalaisissa kielissä. Ilmeisesti ilmaus on levinnyt italiankielisestä ilmauksesta *fare il fiasco*, 'maksaa pullo', joka on ollut korttipelin kierroksen häviäjän kallis velvollisuus.

Huolimatta merkityksenmuutoksen kulttuuri- ja kontekstisidonnaisuudesta on myös merkityksenmuutoksissa paljon säännönmukaisuutta. Jo varhain on esimerkiksi havaittu, että kielen sanaston eräät osat ovat huomattavasti alttiimpia merkityksenmuutoksille kuin toiset. Siinä missä perusliike- ja havaintoverbit (*tulla*, *mennä*, *nähdä*, *kuulla*), säätä tai luonnonilmiöitä kuvaavat sanat (*kivi*, *tuuli*, *vesi*) tai ruumiinosien nimitykset (*käsi*, *jalka*, *pää*) ovat yleensä vanhoja ja merkitykseltään melko muuttumattomia, muuttuvat mielentiloja, teknologioita, raaka-aineita, uskontoa ja muita kulttuuri-ilmiöitä koskevat sanat historiassa suhteellisen nopeasti. Eräisiin tunne-elämyksiin, kuten rakastumiseen ja humalaan, liittyvä sanasto muuttuu jopa yhden sukupolven aikana (*humalassa* → *kännissä* → *tuiterissa* → *kaasussa*).

Keskeisiä merkityksenmuutoksen tyyppejä ovat mm. merkityksen laajeneminen ja merkityksen supistuminen. Suomen kielen *kulkea* on syntynyt sanasta, joka useimmissa uralilaisissa kielissä merkitsee veden virtaamista. Sanan merkitys on siis laajentunut merkitsemään mitä hyvänsä kulkemista. Vastaavasti suomen *kuori*-sanan vastineet etäsukukielissä tarkoittavat lähinnä puun kuorta tai tuohta. Tässäkin näemme merkityksen laajentuvan kaikenlaisiin kuoriin. Yleisempää on kuitenkin merkityksen supistuminen. Esimerkiksi suomen *otsa*-sanan vastineet sukukielissä tarkoittavat 'edessä olevaa'. Suomessa on siis tapahtunut merkityksen supistuminen tarkoittamaan ainoastaan yhtä edessä olevaa asiaa. Vastaavasti suomen *saarnata* näyttää merkinneen alkuaan kaikenlaista puhumista, mikä ilmenee edelleen esimerkiksi karjalassa, inarinsaamessa ja komissa. Sanan merkitys on siis supistunut merkitsemään

yhtä tiettyä puheen tyyppiä. Merkityksen supistumista tapahtuu helposti erityisesti lainautumisen yhteydessä.

Keskeinen merkityksenmuutosten mekanismi on sanan merkityksen metaforistuminen eli kuvaannollinen käyttö. Sellaiset aivan tavalliset sanat kuin *käsittää, tietää* ja *päätellä* ovat metaforaan perustuvia johdoksia sanoista *käsi*, *tie* ja *pää*. Niiden merkitysten voi katsoa olevan jotakin sellaista kuin ’ottaa kädellä kiinni’, ’kulkea tietä pitkin’ ja ’kulkea päähän asti’. Metaforistumista voi usein havaita myös lainasanoissa. Esimerkiksi venäjän sanasta *stradat’* ’kärsiä’ on lainattu karjalan *ruadua* ja suomen *raataa* ’tehdä työtä’. Suomen *känni* taas on saatu alun perin ruotsalaisesta kannua merkitsevästä sanasta.

Metafora on varsin usein taustalla myös kieliopillisen merkityksen syntymisessä. Niinpä esimerkiksi sana *pää* on tuottanut varsin monia kieliopillisia rakenteita. Tien *päässä* ja pöydän *päällä* oleva asia on päähän rinnastettavissa olevassa metaforisessa suhteessa tilaan. Toisaalta silloin kun televisio tai pesukone pannaan päälle, on kyse ruotsin kielestä saadusta syntaktisesta lainasta. Ruotsissa päällä oleva televisio on *på* ”päällä”, joka on ollut metafora. Mahdollisesti kyseessä on ollut alun perin esimerkiksi valon asettaminen pöydälle tai pöydän yläpuolelle, jotta se valaisisi koko tilan. Tästä valon nostamisesta on sitten saatu television ja muiden sähkölaitteiden käynnistämiseen liittyvä päälle panemisen metafora. Metaforisesti käytetään kieliopillistuneissa rakenteissa usein myös sanoja *juuri* (esim. *vuoren juurella*), *ääri* ’reuna’ (*pöydän ääreen*), *käsi* (esim. *asia on johtajan käsissä*), *hammas* (esim. *hän joutui vastustajien hampaisiin*) jne.

## 9. Typologia ja historiallinen kielitiede

Kun kielitypologian tutkimus on 2000-luvulla vauhdittunut, tämä on vaikuttanut myös historiallis-vertailevaan kielentutkimukseen. Monien eri kielten sanastoa kuvaavia typologisia tietokantoja on luotu historialliseenkin tutkimukseen soveltuvaksi materiaaliksi. Esimerkiksi ns.

leksikaalisia assosiaatioita tai kollokaatioita koskeva tutkimus on laajalti selvittänyt, mitkä käsitteet liittyvät toisiinsa maailman kielissä (esim. rakkaus, sydän ja punainen väri, tai lehti, sappi, vihreä väri ja kateus länsimaisessa kulttuuripiirissä). Tällaiset assosiaatiot voivat johtaa merkityksenmuutoksiin sekä toisaalta heijastua sanojen synnyssä johdoksina (List, Greenhill, Anderson, Mayer, Tresoldi & Forkel 2018; Vanhove 2008).

Martin Haspelmathin ja Uri Tadmorin laaja maailman lainasanoja koskeva tutkimus (Haspelmath & Tadmor 2009) osoitti, miten eri käsitepiireihin liittyvä sanasto lainautuu – ja täten myös muuttuu – aivan eri todennäköisyydellä. Tutkimuksessa havaittiin, että nominit ovat helpommin lainautuvia kuin verbit ja että kieliopillista merkitystä kantavien sanojen lainautuminen oli yli kaksi kertaa epätodennäköisempää kuin tavallisen sanaston lainautuminen. Todennäköisimmin lainautuvaksi havaittiin uskontoon ja teknologiaan liittyvä sanasto. Vähiten todennäköisimmin taas lainautui liikkeeseen ja fysikaaliseen todellisuuteen liittyvä sanasto.

**Koleksifikaatioita** eli samalla sanalla nimitettäviä asioita ja tähän liittyvää merkityksenmuutosta koskevaa tietoa on yhdistelty semanttisten muutosten tietokannassa (Bulakh, Ganenkov, Gruntov, Maisak, Rousseau & Zalizniak 2012–2017), johon on koottu eri lähteistä tiedossa olevia semanttisia muutoksia. Tämäntyyppiset tietokannat voivat tulevaisuudessa muuttaa myös merkityksenmuutoksen tutkimuksen lähes samanlaiseksi täsmätutkimukseksi kuin, mitä äänteenmuutosten tutkimus on.

Viime vuosikymmeninä historiallis-vertailevaa metodia on täydennetty myös tietokoneavusteisilla menetelmillä, jotka esimerkiksi laskevat kielten välisiä etäisyyksiä samankaltaisten sanojen avulla (vrt. esim. Syrjänen, Honkola, Korhonen, Lehtinen, Vesakoski & Wahlberg 2013) tai etsivät sanoja, joissa on tapahtunut säännöllisiä muutoksia. Tietokoneavusteisilla ohjelmilla on pyritty myös etsimään perinteisiä kantakieliä suurempien aikasyvyyksien välisiä yhteyksiä eri kielikuntien välillä, mutta läpimurtoa ei alalla ole saavutettu.

# 10. Yhteenveto

Historiallis-vertaileva kielitiede tutkii sanojen, morfeemien ja rakenteiden historiaa vertailemalla nykykieliä ja kirjallisesti tunnettuja vanhoja kielimuotoja. Se selittää nykykielten variaatiota projisoimalla ne menneisyyden vaihtelemattomaan kantakieleen. Kantakielen jakautuminen moniksi kieliksi on monta kertaa historiassa toistunut tapahtuma. Esimerkiksi latina on romaanisten kielten kantakieli. Rekonstruoidut kantakielet ovat historiallisesti todella olemassa olleita kieliä, joista ei kuitenkaan tiedetä kovin paljoa.

Historiallis-vertailevan metodin perusolettama on, että kielen kehitys on äänteellisesti säännöllistä. Myös kielen muotojen ja merkitysten kehityksessä on paljon säännönmukaisuutta, vaikkakaan säännönmukaisuus ei ole aivan yhtä suurta kuin äänteistön kehityksessä.

Historiallis-vertailevan metodin avulla on mahdollista tutkia myös niitä kielen elementtejä, jotka kantakielten asemesta on saatu naapurikielistä lainaamalla. Lainautuminen voi kohdistua sekä sanastoon, muotoihin että rakenteisiin. Metodi pystyy rakentamaan kielten sisäisten ja ulkoisten vaikutusten aikahierarkioita, joita voi kuvata esim. sukupuilla. Sen avulla on mahdollista erottaa toisistaan eri-ikäiset lainavaikutukset ja tätä tietoa on mahdollista edelleen soveltaa esihistorian tutkimukseen.

Historiallis-vertaileva metodi oli 1800-luvun alkupuolelta lähtien kielitieteen keskeinen metodi, mutta 1900-luvun jälkipuoliskolla se muuttui marginaalisemmaksi. Metodia on kuitenkin jatkuvasti täydennetty mm. kielitypologian näkökohtia huomioimalla. Suurinta osaa maailman kielikunnista ei ole kovinkaan syvällisesti tutkittu historiallis-vertailevasta näkökulmasta.

**Aiheesta lisää:**

Anttila, Raimo. 1988. *Historical and Comparative Linguistics*. Amsterdam: John Benjamins.

Campbell, Lyle. 1998. *Historical Linguistics. An Introduction*. Edinburgh: Edinburgh University Press.

Hakulinen, Lauri. 2000 [1941–1946]. *Suomen kielen rakenne ja kehitys*. Helsinki: Helsingin yliopiston suomen kielen laitos.

Hock, Hans & Joseph, Brian D. 2009. *Language History, Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics*. Berlin: Mouton de Gruyter.

Häkkinen, Kaisa. 1990. *Mistä sanat tulevat. Suomalaista etymologiaa*. Tietolipas 117. Helsinki: Suomalaisen Kirjallisuuden Seura.

Lehtinen, Tapani. 2007. *Kielen vuosituhannet. Suomen kielen kehitys varhaissuomesta kantauraliin*. Tietolipas 215. Helsinki: Suomalaisen Kirjallisuuden Seura.

## KIRJALLISUUS


Aikio, Ante. 2008. *Saami Borrowings in Finnish and Karelian*. Oulu: Oulun yliopisto.

Anttila, Raimo. 1988. *Historical and Comparative Linguistics*. Amsterdam: John Benjamins.

Bulakh, Maria, Ganenkov, Dmitry, Gruntov, Ilya, Maisak, Timur, Rousseau, Maxim & Zalizniak, Anna. 2012–2017. *Database of Semantic Shifts in the Languages of the World*. [verkkoaineisto]. [viitattu 2.5.2018]. Saatavissa: http://semshifts.iling-ran.ru/.

Campbell, Lyle. 1998. *Historical Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

Fox, Anthony. 1995. *Linguistic Reconstruction. An Introduction to Theory and Method*. Oxford: Oxford University Press.

Haspelmath, Martin & Tadmor, Uri. 2009. *Loanwords in the World's Languages*. New York (NY): Mouton de Gruyter.

Heine, Bernd & Kuteva, Tania. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.

Hock, Hans Heinrich & Joseph, Brian D. 2009. *Language History, Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics*. Berlin: Mouton de Gruyter.

Häkkinen, Jaakko. 2007. Uralilaisen kantakielen ajoitus ja paikannus. Perustelut puntarissa. *Journal de la Société Finno-Ougrienne* 91, 9–56.

Häkkinen, Kaisa. 1990. *Mistä sanat tulevat? Suomalaista etymologiaa*. Tietolipas 117. Helsinki: Suomalaisen Kirjallisuuden Seura.

Häkkinen, Kaisa. 2013. *Nykysuomen etymologinen sanakirja*. Helsinki: SanomaPro.

Idström, Anna, Piirainen, Elisabeth & Falzett, Tiber (toim.). 2012. *Endangered Metaphors. Cognitive Linguistic Studies in Cultural Contexts*. Vol. 2. Amsterdam: John Benjamins.

Juvonen, Päivi & Koptjevskaja-Tamm, Maria. 2016. *The Lexical Typology of Semantic Shifts*. Cognitive Linguistic Research 58. Berlin: De Gruyter Mouton.

Kallio, Petri. 2012. The non-initial syllable vowel reductions in Finnic. Julkaisussa: Hyytiäinen, Tiina, Jalava, Lotta, Saarikivi, Janne & Sandman, Erika (toim.) *Per Urales ad Orientem. Iter polyphonicum multilingue. Festskrift tillägnad Juha Janhunen på hans sextioårsdag den 12 februari 2012*. Mémoires de la Société Finno-Ougrienne 264. Helsinki: Suomalais-Ugrilainen Seura, 163–175.

Koivulehto, Jorma. 1991. *Uralische Evidenz für die Laryngaltheorie*. Veröffentlichungen der Kommission für Linguistik und Kommunikationsforschung 24. Sitzungsberichte der Österreichishe Akademie der Wissenschaften, Philosophisch-historische Klasse. Bd 566. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Kortelainen, Kristiina. 2012. Ei ole kaikki muumit laaksossa. Tutkimus suomen kielen idiomikonstruktion produktiivisuudesta. Julkaisematon pro gradu -tutkielma. Kieli- ja käännöstieteiden laitos, Tampereen yliopisto. Saatavissa: http://urn.fi/URN:NBN:fi-fe201201301267.

Kulonen, Ulla-Maija. 2002. Kielitiede ja suomen väestön juuret. Julkaisussa: Grünthal, Riho (toim.) *Ennen, muinoin. Miten menneisyyttämme tutkitaan*. Tietolipas 180: Helsinki: Suomalaisen Kirjallisuuden Seura, 102–116.

Lehtinen, Tapani. 2007. *Kielen vuosituhannet. Suomen kielen kehitys kantauralista varhaissuomeen*. Tietolipas 215. Helsinki: Suomalaisen Kirjallisuuden Seura.

List, Johann-Mattis, Greenhill, Simon, Anderson, Cormac, Mayer, Thomas, Tresoldi, Tiago & Forkel, Robert (toim.). 2018. Database of Cross-Linguistic Colexifications. Jena: Max Planck Institute for the Science of Human History. [verkkoaineisto]. [viitattu 29.5.2018]. Saatavissa: http://clics.clld.org.

Lüpke, Friederike. 2013. *Repertoires and Choices in African Languages*. Boston: de Gruyter Mouton.

Mikkola, Jooseppi J. 1894. *Berührungen zwischen den westfinnischen und slavischen Sprachen*. I. *Slavische Lehnwörter in den westfinnischen Sprachen*. SUST VIII. Helsingfors: Druckerei der finnischen Litteratur-Gesellschaft.

Milroy, James. 1992. *Linguistic Variation and Change: On the Historical Sociolinguistics of English*. Oxford: Blackwell.

Parpola, Simo. 2016. *Etymological Dictionary of the Sumerian language*. Publications of the Foundation of Finnish Assyriological Research, Vol. 16. Winona Lake: Eisenbrauns.

Paul, Hermann. 1881. *Prinzipien der Sprachgeschichte*. Halle: Max Niemeyer.

Ravila, Paavo. 1946. *Johdatus kielihistoriaan*. Tietolipas 3. Helsinki: Suomalaisen Kirjallisuuden Seura.

Saarikivi, Janne. 2006. *Substrata Uralica. Studies on Finno-Ugrian Substrate in Northern Russian dialects*. Tartu: Tartu University Press. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-4519-6.
Saarikivi, Janne. 2016. Sumerisch ist weiterhin eine isolierte Sprache [Rez. von Simo Parpola: Etymological dictionary of the Sumerian language]. *Finnisch-Ugrische Forschungen* 63. Helsinki: Finnisch-Ugrische Gesellschaft.
Saarikivi, Janne. (painossa). Diversification of Proto-Uralic. Julkaisussa: Bakró-Nagy, Marianne, Laakso, Johanna & Skribnik, Elena (toim.) *The Oxford Guide to the Uralic Languages*. Oxford: Oxford University Press.
Sajnóvics, János. 1964. *Demonstratio. Idioma ungarorum et lapponum idem esse*. (Alun perin julkaistu 1770). Indiana University Publications, Uralic and Altaic Series 91. Bloomington: Indiana University Press.
Salminen, Tapani. 1999. Euroopan kielet muinoin ja nykyisin. Julkaisussa: Fobelgerb, Paul (toim.) *Pohjan poluilla: suomalaisten juuret nykytutkimuksen mukaan*. Bidrag till kännedom av Finlands natur och folk 153. Helsinki: Finska vetenskaps-societeten, 13–26.
Schiffman, Harold F. 1996. *Linguistic Culture and Language Policy*. London & New York (NY): Routledge.
Sirola-Belliard, Maija. 2016. *Suomen komitatiivi* -ine *ja* kanssa. Tampere: Tampere University Press.
SKES = Toivonen, Yrjö H., Itkonen, Erkki & Joki, Aulis J. 1955–1978. *Suomen kielen etymologinen sanakirja* 1–7. Lexica Societatis Fenno-Ugricae 12. Helsinki: Suomalais-Ugrilainen Seura.
SSA = Itkonen, Erkki & Kulonen, Ulla-Maija (toim.). 1992–2000. *Suomen sanojen alkuperä 1–3. Etymologinen sanakirja*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 62. Suomalaisen Kirjallisuuden Seuran Toimituksia 556. Helsinki: Suomalaisen Kirjallisuuden Seura & Kotimaisten kielten tutkimuskeskus.
Syrjänen, Kai, Honkola, Terhi, Korhonen, Kalle, Lehtinen, Jyri, Vesakoski, Outi & Wahlberg, Niklas. 2013. Shedding more light on language classification using basic vocabularies and phylogenetic methods. A case study of Uralic. *Diachronica* 30, 323–352.
Thomsen, Vilhelm. 1870. Über den einfluss der germanischen sprachen auf die finnischlappischen: eine sprachgeschichtliche untersuchung. Halle: Verlag der Buchhandlung des Waisenhauses.
Toivanen, Reetta & Saarikivi, Janne (toim.). 2016. *Linguistic Genocide or Superdiversity. New and Old Language Diversities*. Linguistic Diversity and Language Rights, Vol. 14. Bristol: Multilingual Matters.
Vanhove, Martine. 2008. *From Polysemy to Semantic Change. Towards a Typology of Lexical Semantic Associations*. Amsterdam: John Benjamins.

# Luku 16 Antropologinen lingvistiikka

*Ekaterina Gruzdeva*
https://orcid.org/0000-0001-8898-649X
*Juha Janhunen*
https://orcid.org/0000-0003-1242-8808

**Mikä?**

*Antropologia* tieteenalan nimityksenä tarkoittaa sananmukaisesti ”oppia” (kreikan *-logía*) ”ihmisestä” (kreikan *ánthrôpos*). Antropologinen lingvistiikka on siis kielitieteen osa-alue, joka on kiinnostunut kielen kulttuurisesta ja sosiaalisesta merkityksestä. Siinä yhdistyvät kielitieteen, kulttuurientutkimuksen ja yhteiskuntatieteiden menetelmät.

**Katso myös:**

Luku 5 Lingvistinen etnografia toimintayhteisöissä

Luku 13 Kielen dokumentointi ja kieliopin kuvaus

Luku 14 Kielitypologian menetelmät

Luku 15 Historiallinen ja vertaileva kielentutkimus

Luku 21 Sosiolingvistiikka

# 1. Johdanto: mitä antropologinen lingvistiikka on?

*Antropologia* tieteenalan nimityksenä tarkoittaa sananmukaisesti ”oppia” (kr. -*logía*) ”ihmisestä” (kr. *ánthrôpos*). Ihmistä voidaan tutkia fyysisenä lajina (fyysinen antropologia), kulttuuria luovana ja kantavana olentona (kulttuuriantropologia) tai myös yhteisöjä rakentavana ja ylläpitävänä ryhmän jäsenenä (sosiaaliantropologia). Koska ihmisen tärkein ominaisuus sekä kulttuurin että yhteisöllisyyden kannalta on kieli, myös kieltä voidaan tutkia antropologian kannalta ja antropologian menetelmin. Tätä tutkimusta tekee antropologinen lingvistiikka.

Antropologinen lingvistiikka on siis kielitieteen osa-alue, joka on kiinnostunut kielen kulttuurisesta ja sosiaalisesta merkityksestä. Kulttuuriset ja sosiaaliset ilmiöt näkyvät aina sekä kielen substanssissa että rakenteessa – sanastossa ja kieliopissa – mutta myös kielen kontekstissa – kieliyhteisön rakenteessa, historiassa ja kontakteissa muiden kieliyhteisöjen kanssa. Aika ja paikka ovat keskeisiä kielen antropologisen kontekstin ymmärtämiseksi. Hyvin erityyppisen näkökulman kieleen verrattuna antropologiseen lingvistiikkaan muodostavat esimerkiksi erilaiset kielen formaalit teoriat, joiden avulla kieliä voidaan kuvata järjestelminä riippumatta niiden kontekstista.

Koska kulttuurit ja ihmisyhteisöt voivat olla keskenään hyvin erilaisia, antropologisessa lingvistiikassa korostuu kielten erilaisuus eli diversiteetti. Maapallolla puhutaan tällä hetkellä noin 6 000–7 000 kieltä, ja antropologinen lingvistiikka kohdistaa huomionsa erityisesti niihin maailman kieliin, joita ei vielä ole kuvattu tai edes luokiteltu ja jotka lisäksi ovat nopeassa tahdissa katoamassa kolonialismin ja globalisaation aiheuttamassa yleismaailmallisessa murroksessa. Vaikka antropologinen lingvistiikka voi hyödyntää yleisten kieliteorioiden ja kielitypologian luomia malleja, se pyrkii näkemään näiden taakse ja löytämään ne kontekstuaaliset ominaisuudet, jotka tekevät kustakin kielestä ainutkertaisen.

Antropologinen lingvistiikka tutkii myös sitä, kuinka kieltä käytetään suhteessa kielten sisäiseen ja ulkoiseen variaatioon. Hyvin suuri osa ihmiskunnasta käyttää päivittäin useita kieliä tai kielimuotoja – murteita,

tyylejä ja rekistereitä –, ja tämän moninaisuuden takana on monenlaisia sosiaalisia, esimerkiksi etnisyyteen, sukupuoleen, kulttuuriin ja yhteiskuntarakenteisiin liittyviä tekijöitä. Näitä tekijöitä eivät formaalit kieliteoriat tai kognitiivinen lingvistiikka yleensä huomioi, mutta ne kuuluvat erityisesti antropologisen lingvistiikan kompetenssiin. Tämä merkitsee samalla sitä, että antropologinen lingvistiikka on luonteeltaan monitieteistä ja tieteidenvälistä tutkimusta: siinä yhdistyvät kielitieteen, kulttuurientutkimuksen ja yhteiskuntatieteiden menetelmät.

Antropologinen lingvistiikka ei ole kuitenkaan alana täysin yhtenäinen, vaan sen sisällä on erilaisia painotuksia, ja myös sen ulkoiset rajat muihin tieteisiin ovat liukuvat. Näitä eroja ja rajoja heijastavat ne erilaiset termit, joilla tutkijat operoivat. Seuraavassa tarkastellaan antropologisen lingvistiikan suhdetta joihinkin lähialoihin.

**Antropologinen lingvistiikka ja lingvistinen antropologia**. *Antropologisen lingvistiikan* rinnalla puhutaan usein *lingvistisestä antropologiasta*. Vaikka näitä termejä voidaan käyttää samassa merkityksessä, niiden välillä voidaan nähdä myös painotusero. Antropologille kieli on ennen kaikkea väline, jonka kautta voidaan lähestyä kulttuurin ja yhteiskunnan rakenteita, kun taas lingvistille kieli on varsinainen tutkimuskohde, jossa näkyvät ympäröivän kulttuurin ja yhteiskunnan rakenteet. Joidenkin tutkijoiden mielestä on ylipäänsä turha puhua lingvistisestä antropologiasta, koska kieli ihmisen lajiominaisuutena kuuluu aina antropologian piiriin. Toiset taas pitävät myös antropologista lingvistiikkaa ensisijaisesti antropologiana, koska erityisesti anglosaksista lingvistiikkaa ovat viime vuosikymmeninä dominoineet antropologialle vieraat universalistiset ja formalistiset näkemykset.

Selvästi antropologian piiriin kuuluvat joka tapauksessa ne kysymykset, jotka liittyvät kieleen ihmisen lajiominaisuutena. Nykyään puhuttavat kielet antavat vain hyvin kapean näkökulman ihmiskielen esihistorialliseen taustaan. Kieli on ollut olemassa ehkä jo satoja tuhansia vuosia, ja sen fyysistä alkuperää voi tutkia lähinnä vain fossiililöydöistä tehtyjen johtopäätösten avulla (paleoantropologia). Arviot muinaisesta kielidiversiteetistä edellyttävät myös ennen kaikkea antropologista tietoa ihmisyhteisöjen koosta, sijainnista ja elintavoista. Varsinainen antropologinen lingvistiikka liikkuu yleensä lähempänä nykyaikaa tai

suorastaan nykyajassa, jolloin kieltä voidaan havainnoida sen välittömässä kulttuurisessa ja sosiaalisessa kontekstissa.

**Antropologinen lingvistiikka ja etnolingvistiikka**. Antropologisen lingvistiikan synonyyminä käytetään usein myös termiä *etnolingvistiikka*. Etnolingvistiikka viittaa sananmukaisesti etnisyyteen ja etnisiin ryhmiin (kr. *éthnos*), ja koska yksi antropologisen lingvistiikan painopistealueista on tyypillisesti pienten etnisten ryhmien aikaisemmin tutkimattomien kielten kartoitus ja dokumentaatio, etnisyyteen viittaaminen on paikallaan. Kyse on kuitenkin myös terminologisten traditioiden erosta: etnisten ryhmien tutkimus tunnetaan varsinkin Euroopassa perinteisesti ”etnologiana”, kun taas Pohjois-Amerikassa puhutaan useammin ”antropologiasta” (tarkoittaen erityisesti kulttuuriantropologiaa).

On kuitenkin hyvä muistaa, että sekä lingvistinen antropologia että antropologinen lingvistiikka voivat tutkia myös muita kuin kielen etnisiä ulottuvuuksia. Esimerkiksi suurkaupunkien sosiaalisen diversiteetin tutkimuksesta puhutaan usein urbaanina antropologiana, ja myös antropologinen lingvistiikka voi tutkia kielten kontekstuaalisia ulottuvuuksia etnisyydestä riippumatta. Tältä kannalta etnolingvistiikka on parhaiten ymmärrettävissä antropologisen lingvistiikan alalajina, jossa huomio kiinnitetään nimenomaan kielen ja etnisyyden yhteyksiin. Tämä on sinänsä tärkeä yhteys, koska kieli on kiistämättä tärkein etnistä identiteettiä ylläpitävä tekijä.

**Antropologinen lingvistiikka ja sosiolingvistiikka**. Koska antropologinen lingvistiikka tarkastelee kieltä suhteessa kieliyhteisöön ja sen sosiaalisiin rakenteisiin, se on tutkimusalana lähellä myös sosiolingvistiikkaa. Kieliyhteisön koko ja sosiaalinen koostumus, muun muassa puhujien ikä- ja elinkeinorakenne sekä liikkuvuus, vaikuttavat olennaisella tavalla siihen, miten kieli elää ja toimii ympäristössään, ja tämän vaikutuksen tutkimus kuuluu luonnostaan antropologisen lingvistiikan piiriin. Enemmän kuin ”puhtaassa” sosiolingvistiikassa antropologisessa lingvistiikassa korostuvat kuitenkin kielen sosiaalisen variaation kulttuuriset ja etniset taustatekijät.

Sosiolingvistiikan kehittämiä metodeja voidaan joka tapauksessa menestyksellisesti hyödyntää antropologisessa lingvistiikassa, eikä jyrkkää rajaa näiden lähestymistapojen välille voida vetää (**ks. Sosiolingvistiikka**

tässä kirjassa). Antropologisen lingvistiikan sosiolingvistiset ulottuvuudet sivuavat jossain määrin myös psykolingvistiikkaa, mutta tutkimusalana psykolingvistiikka on kauempana antropologisesta näkökulmasta ja kuuluu enemmän kognitiivisen kielentutkimuksen kenttään. Psykolingvistiikka tutkii ennen kaikkea kielen suhdetta kognitiivisiin prosesseihin (ajatteluun), eivätkä kielen kulttuuriset ja sosiaaliset ulottuvuudet tyypillisesti ole mukana sen kysymyksenasetteluissa.

**Antropologinen lingvistiikka ja ekolingvistiikka**. Vasta varsin hiljattain antropologisen lingvistiikan rinnalle on kehittynyt ekolingvistiikkana tunnettu tutkimussuuntaus, jota on kehittänyt mm. Australiassa vaikuttava saksalainen lingvisti Peter Mühlhäusler (1996). Ekolingvistiikka eli lingvistinen ekologia ottaa keskeisen näkökulmansa ekologiasta, joka tutkii erityisesti luonnonympäristön rakenteita. Ekolingvistiikka on alana lähellä sosiolingvistiikkaa mutta eroaa siitä muun muassa sikäli, että se ottaa kielen sosiaalisen ympäristön lisäksi huomioon myös fyysiset ympäristötekijät, esimerkiksi ilmaston ja kasvillisuuden, joilla puolestaan on suora vaikutus kieliyhteisön kulttuuriin, elinkeinorakenteeseen ja kokoon.

Ekolingvistiikka tarjoaa antropologiselle lingvistiikalle erityisen käyttökelpoisia malleja kielten historiallisen leviämisen, jakautumisen, vuorovaikutuksen ja viime kädessä uhanalaisuuden tutkimukseen. Kielet muodostavat, kuten elävän luonnon lajit, ekologisia järjestelmiä, joiden pysyvyydelle on olennaista tasapainon säilyminen. Ekolingvistiikka, yhtä vähän kuin antropologinen lingvistiikkakaan, ei kuitenkaan välttämättä ota kantaa tasapainon säilymisen puolesta, vaan se voi yhtä hyvin tutkia niitä tilanteita, joissa tasapaino horjuu.

## 2. Antropologisen lingvistiikan taustaa

Antropologinen lingvistiikka on kehittynyt yhdessä muun kielitieteen kanssa, ja on vaikea sanoa, missä sen varsinaiset juuret ovat. Voidaan jopa ajatella, että lähes kaikki sellainen kielentutkimus, joka hakee

aineistonsa puhutuista kielistä, on antropologista lingvistiikkaa riippumatta siitä, millä nimellä tällaista tutkimusta kulloinkin on kutsuttu. Puhuttuja kieliä taas on ainakin eurooppalaisessa kontekstissa tutkittu jo vähintään löytöretkien ja reformaation ajoista alkaen.

**Amerikkalainen antropologia**. Usein antropologisen lingvistiikan katsotaan olevan lähtöisin 1800-luvun lopun pohjoisamerikkalaisesta antropologiasta, joka alkoi tutkia intiaanikansoja ja niiden kieliä – samalla kun intiaaniheimoja systemaattisesti tuhottiin Yhdysvaltain laajentuessa kohti länttä. Tämän tutkimuksen pioneerina ja koko antropologian pohjoisamerikkalaisen koulukunnan perustajana tunnetaan Franz Boas (1858–1942), jonka johdolla valmistui muun muassa massiivinen teossarja *Handbook of American Indian Languages* (Boas 1911–1922). Boasin koulukuntaan liittyivät myös venäläiset Vladimir Bogoraz (1865–1936) ja Vladimir Jochelson (1855–1937), joiden tutkimuskohteena olivat Koillis-Siperian kansat ja kielet.

Boasin koulukunta korosti alusta alkaen sitä, ettei ole olemassa ”primitiivisiä” kieliä tai kansoja, vaan kaikki kielet ja kulttuurit ovat omalla tavallaan yhtä kompleksisia. Samalla ymmärrettiin, että kaikki kielet ja kulttuurit ovat keskenään hyvin erilaisia, ja tämän erilaisuuden kartoittamiseksi ne tuli perusteellisesti dokumentoida. Dokumentaatio tapahtui kenttätyön avulla välittömässä kontaktissa kielen puhujiin, ja kenttätyötä pyrittiin tekemään erityisesti sellaisilla alueilla ja sellaisten kielten parissa, joita vielä ei ollut kuvattu. Näitä alueita olivat Pohjois-Amerikan lisäksi erityisesti Keski- ja Etelä-Amerikka mutta myös Australia, Oseania, Afrikka sekä suuret osat Aasiaa – alueita, jotka kaikki ovat nykyäänkin antropologisen lingvistiikan keskeisiä kohteita.

Boasin koulukunnalle oli – ja on – ominaista myös neopositivistinen käsitys siitä, että empiirisesti ”aitoa” kieliainesta on ainoastaan se, joka on kielenpuhujan spontaanisti tuottamaa ja puhujalta suoraan dokumentoitua. Tältä pohjalta tutkimus keskittyi erityisesti tekstimuotoisten kielennäytteiden keräämiseen. Teksteistä voitiin poimia esimerkkejä sekä sanaston että kieliopin tutkimukseen, ja niitä voitiin käyttää myös lähteinä kulttuuria lähestyttäessä. Ylipäänsä kielen ja kulttuurin tutkimusta ei boasilaisessa käsityksessä eroteta toisistaan, koska tutkimuksessa lähdetään siitä, että kulttuuria voi tutkia vain kielen kautta ja että

kielessä joka tapauksessa heijastuvat puhujien edustaman kulttuurin ominaisuudet.

**Eurooppalainen etnologia**. On kuitenkin virhe pitää antropologista lingvistiikkaa yksinomaan pohjoisamerikkalaisena ”keksintönä”, sillä ala on ollut pitkään tunnettu ja harrastettu myös Euroopassa. Boasin eurooppalaisena vastineena pidetään usein Englannissa vaikuttanutta puolalaista Bronisław Malinowskia (1884–1942), jonka kohdealueena olivat Tyynenmeren saaret, mutta hänen rinnallaan on mainittava niin ikään Englannissa uransa luonut suomenruotsalainen Edward Westermarck (1862–1939), joka teki kenttätyönsä Pohjois-Afrikassa. Malinowski ja Westermarck kuuluvat eittämättä modernin etnologian perustajahahmoihin, mutta myös heillä oli merkittävä edeltäjä, Australiassa toiminut venäläinen Nikolai Miklukho-Maklai (1846–1888), joka yhtenä ensimmäisistä kartoitti kentällä Tyynenmeren kansoja.

On tosin jossain määrin kyseenalaista, voidaanko sen paremmin pohjoisamerikkalaisia antropologeja kuin eurooppalaisia etnologejakaan pitää nimenomaan lingvistisen antropologian edelläkävijöinä, sillä heille kaikille kieli oli ensi sijassa vain väline, jonka kautta päästiin tutkimaan kulttuurin syvempiä rakenteita. Sinänsä on selvää, että kulttuuria tai sosiaalisia rakenteita ei voi tutkia ilman niitä kannattavan yhteisön kieltä, mutta kielentutkimus eroaa kulttuurientutkimuksesta juuri siinä, että sille kieli ei ole väline vaan varsinainen tutkimuskohde. Voi sanoa, että kulttuurientutkimus on parhaimmillaankin vain välineellistettyä kielentutkimusta.

**Kielitieteilijät lähetystyössä.** Eräänlaista välineellistettyä kielentutkimusta edustaa myös se traditio, jonka perimmäisenä tarkoituksena on uskonnon levittäminen ja sen kautta kulttuurisen yhdenmukaisuuden edistäminen. Kristilliset lähetyssaarnaajat ovat jo vuosisatoja, aluksi kolonialismin ja nyttemmin globalisaation hengessä, kartoittaneet maailman kielellistä moninaisuutta voidakseen levittää sanomaansa myös sellaisille kansoille, joiden kielillä ei ole kirjallista perinnettä. Tätä toimintaa ovat harjoittaneet kaikki kristilliset suuntaukset, mutta erityisen vahvaksi sen asema on muodostunut Pohjois-Amerikassa syntyneiden protestanttisten liikkeiden ja vapaakirkollisten suuntausten keskuudessa. Esimerkiksi suomalaisen afrikanistiikan perinne on syntynyt lähes yksinomaan tältä pohjalta.

Uskonnollispohjaista kielentutkimusta ei ole syytä vähätellä, sillä se on vuosikymmenien ajan, erityisesti juuri Pohjois-Amerikassa, ylläpitänyt antropologisen lingvistiikan traditiota, samalla kun muu kielitiede on painottunut voimakkaammin universalistisen ja kognitiivisen tutkimuksen suuntaan. Valtaosa aikaisemmin tutkimattomien kielten kuvauksista syntyy nykyäänkin kristillisten yhteisöjen tuella ja rahoituksella. Keskeinen rooli tässä toiminnassa on ”Kielitieteen kesäkoululla” (Summer Institute of Linguistics eli SIL International), joka systemaattisesti kouluttaa kielitieteellisesti päteviä lähetystyöntekijöitä eri puolille maailmaa. Tunnetuin tämän koulun kasvateista oli Kenneth L. Pike (1912–2000), jota voi perustellusti pitää yhtenä antropologisen lingvistiikan suurista nimistä.

**Suomalainen etnolingvistiikka**. On monia syitä korostaa Suomen merkitystä antropologisen lingvistiikan historiassa. Suomalaisilla tutkijoilla oli Venäjän yhteydessä poikkeuksellisen hyvät mahdollisuudet tutkia Venäjän ja Siperian kielellistä moninaisuutta, ja tähän oli myös hyvät syyt, sillä haluttiin löytää suomalaisten kielisukulaiset ja määrittää ”Suomen suvun” rajat. Tästä tutkimusperinteestä, jonka keskeisenä metodina oli kielitieteellinen kenttätyö, nousivat Suomen ”kansalliset kielitieteet” uralistiikka (fennougristiikka) sekä myös altaistiikka, jotka myöhemmin levisivät maailmalle ja tulivat osaksi kansainvälistä kielentutkimusta.

Suomalaisen etnolingvistiikan perustajina voi pitää Anders Johan Sjögreniä (1794–1855) ja Matthias Alexander Castrénia (1813–1852), joiden tutkimat kielet yhdessä kattoivat suurimman osan Venäjän imperiumia Kaukasiaa ja Siperiaa myöten. Tätä perinnettä jatkoi 1883 perustetun Suomalais-Ugrilaisen Seuran tuella suuri joukko suomalaisia kielentutkijoita, jotka etnolingvistiikan hengessä dokumentoivat kentällä sekä kieliä että niiden konteksteja. Tyypillistä tälle vaiheelle oli, että kukin tutkija yleensä erikoistui vain yhteen kieleen tai kielihaaraan, kuten Artturi Kannisto (1874–1943) mansiin, K. F. Karjalainen (1871–1919) hantiin ja Toivo Lehtisalo (1887–1962) nenetsiin. Monipuolisempaa lähestymistapaa edustivat Kai Donner (1888–1935) ja G. J. Ramstedt (1873–1950), joista viimeksi mainittu tunnetaan nimenomaan modernin altailaisen kielentutkimuksen perustajana.

Suomalainen etnolingvistiikka syntyi noin puoli vuosisataa ennen amerikkalaista antropologiaa, mutta koska sen tulokset julkaistiin

aikoinaan pääasiassa ruotsiksi ja saksaksi, se on tieteenhistoriassa jäänyt angloamerikkalaisen tutkimuksen varjoon. Mielenkiintoista on, että myös suomalaiset tutkijat siirtyivät – Boasin koulukunnasta riippumatta – kohti neopositivistista empirismiä ja keskittyivät yhä enemmän tekstien keruuseen ja yhä vähemmän kielen rakenteen kuvaukseen. Alan uranuurtajat Sjögren ja Castrén käyttivät vielä runsaasti myös elisitaatiomenetelmiä, minkä ansiosta he saattoivat kuvata kielioppia järjestelmällisemmin ja täydellisemmin kuin pelkän tekstiaineiston varassa toimivat seuraajansa.

**Saksalainen humanismi**. Suomalaisen etnolingvistiikan juuret puolestaan olivat saksalaisen humanismin pohjalta syntyneissä kielitieteen suuntauksissa, jotka johtivat historiallis-vertailevan kielentutkimuksen ja kielitypologian syntyyn (ks. Historiallinen ja vertaileva kielentutkimus; Kielitypologian menetelmät tk.). Tosin historiallis-vertaileva kielentutkimuskin mainitaan usein harhaanjohtavasti anglosaksien keksintönä William Jonesin (1746–1794) intialaisten ja eurooppalaisten kielten sukulaisuutta koskevan lyhyen kannanoton johdosta, mutta alan varsinainen perustaja oli saksalainen sanskritisti Franz Bopp (1791–1867), joka 1816 julkaisi systemaattisen esityksen sanskritin, persian, kreikan, latinan ja germaanisten kielten yhteisistä piirteistä (Bopp 1816). Tämä indoeurooppalaisen kielentutkimuksen perusteos toimi lähtökohtana muiden kielikuntien, myös uralilaisten kielten, vertailevalle tutkimukselle.

Saksalaiseen humanismiin pohjautui viime kädessä myös tanskalaisen Rasmus Raskin (1787–1832) työ. Rask kirjoitti suuren joukon kielioppeja sekä tunnetuista että vähemmän tunnetuista eurooppalaisista kielistä, muun muassa useista germaanisista ja romaanisista kielistä sekä pohjoissaamesta. Hän teki myös matkan Intiaan ja julkaisi kenttätyöhön perustuvia havaintoja joukosta intialaisia kieliä. Raskilla oli merkittävä rooli suomalaisen etnolingvistiikan historiassa, sillä hänen vierailunsa Suomessa 1818 innoitti Sjögreniä antautumaan kielitieteilijän uralle. Myöhemmin hänen elämäntyönsä innoitti myös Castrénia. Kenttälingvisteinä sekä Sjögren että erityisesti Castrén olivat kuitenkin Raskia huomattavasti merkittävämpiä.

Jo ennen vertailevan kielentutkimuksen metodien muotoutumista saksalaisen humanismin piiristä nousi kielitieteen kannaltakin

merkittäviä polyhistoreita ja filosofeja, kuten Johann Gottfried von Herder (1744–1803) ja Wilhelm von Humboldt (1767–1835). Teoksessaan kielen alkuperästä Herder (1772) painotti erityisesti kielen merkitystä ajattelun ja sen kautta kulttuurin perustana. Tästä oli vain lyhyt matka Humboldtin näkemykseen kielestä kansallisen identiteetin keskeisenä tunnusmerkkinä, mikä puolestaan oli lähtökohtana suomalaiselle etnolingvistiikalle, joka kielen kautta etsi suomalaisuuden juuria. Humboldt korosti myös kielidiversiteetin tutkimuksen tärkeyttä. Hänen postuumisti 1836 julkaistu johdatuksensa Jaavalla puhuttuun kawin kieleen on yksi antropologisen lingvistiikan klassikoista.

**Valistusajan kenttätyöperinne**. Kielitieteellisen kenttätyön perinne ja kielidiversiteetin tutkimus saivat kuitenkin alkunsa jo valistusajalla, jolloin alettiin systemaattisesti kerätä aineistoa vähän tunnetuista kielistä eri puolilta maailmaa. Tälläkin alalla tutkimusta johtivat saksalaiset polyhistorit, joista monet olivat Venäjän tiedeakatemian palveluksessa ja keräsivät aineistoa Venäjän keisarikunnan kielistä eli lähes koko Keski- ja Pohjois-Euraasiasta. Kielet olivat tyypillisesti vain yksi dokumentaatiokohde monien muiden joukossa, sillä samalla kuvattiin myös kulttuureja ja historiaa sekä luonnonympäristöä – muun muassa kasveja, eläimiä ja mineraaleja. Etnobotaniikka, joka tutkii kasveja suhteessa niitä käyttävien väestöjen kulttuureihin ja kieliin, on esimerkki tutkimusaloista, jotka versoivat valistusajan kenttätyöperinteestä.

Varsinaisina kielitieteellisinä tuloksina valistusaika dokumentoi lähinnä sanastoa, harvoin kokonaisia fraaseja eikä juuri koskaan kielioppia. Varhaisen pioneerityön tekivät Venäjän Kiinan-lähettiläänä toiminut tanskalainen Eberhard Isbrand Ides (1657–1708), Venäjällä neuvonantajana toiminut hollantilainen Nicholaus Witsen (1641–1717) sekä Venäjälle sotavankina joutunut ruotsalainen Philip Johan von Strahlenberg (1676–1747), joiden kaikkien kuvaukset Pohjois- ja Keski-Aasiasta sisältävät myös kieliaineistoa. Ensimmäisen systemaattisen ”koko maailman” kielten sanastokokoelman julkaisi Peter Simon Pallas (1741–1811), ja vastaavanlaisia sanastoja ilmestyi myöhemmin muitakin, muun muassa Julius Klaprothilta (1783–1835). Näissä sanastoissa kieliä ryhmiteltiin sekä geneettisesti että alueellisesti, ja tältä pohjalta alkoi vähitellen kirkastua kuva maailman kielikunnista ja myös kielialueista.

## 3. Antropologisen lingvistiikan menetelmiä

Koska antropologisen lingvistiikan kohteena ovat sekä kielet että niiden kulttuuriset, sosiaaliset, alueelliset ja historialliset kontekstit, ala on perusluonteeltaan monitieteinen. Lähtökohtana on kuitenkin aina kulloinkin tutkittava kieli tai kielimuoto, jonka kielitieteellinen kuvaus on edellytyksenä sen kontekstin tutkimiselle. Kun sekä kieli että sen kontekstuaaliset ulottuvuudet on monipuolisesti kartoitettu, voidaan edetä analyysin tasolle ja tehdä esimerkiksi johtopäätöksiä kielen ja kieliyhteisön historiasta sekä vertailuja eri kielten ja kieliyhteisöjen välillä.

**Kielitieteellinen kenttätyö**. Lähes kaikki antropologinen lingvistiikka perustuu kenttätyön kautta hankittuun aineistoon ja sen analyysiin. Kenttätyön perusmenetelmät ovat samat riippumatta siitä, tutkitaanko kieltä vai sen kontekstia: tutkittava kohde dokumentoidaan mahdollisimman perusteellisesti havainnoinnin ja haastattelujen kautta (ks. Kielten dokumentointi ja kieliopin kuvaus tk.). Kohderyhmänä on yleensä kieliyhteisö, jonka parissa lingvisti viettää riittävästi aikaa voidakseen suorittaa dokumentoinnin monipuolisesti ottaen huomioon yhteisön sisällä vallitsevat sosiaaliset erot esimerkiksi sukupuolten ja ikäryhmien välillä. Kielen jonkinasteiseen dokumentointiin riittää kuitenkin hätätilassa yksikin puhuja, ja varsin monia kieliä onkin dokumentoitu vain yhden – usein viimeisen – puhujan avulla. Pelkän kielen dokumentointi on mahdollista myös laboratorioympäristössä, mutta kieliyhteisön kokonaisprofiilin voi kartoittaa vain kentällä yhteistyössä kielenpuhujien kanssa.

Koska vähänkään isomman kieliyhteisön kaikkia puhujia ei voida tavoittaa, lingvisti käyttää tyypillisesti apunaan pientä joukkoa kielenoppaita (informantteja eli konsultantteja). Kielenoppaiden valinta on ensimmäinen tehtävä kenttätyötä aloitettaessa, ja se on syytä tehdä huolellisesti ottaen huomioon myös sen, että eri kielenoppaat saattavat soveltua erilaisiin tehtäviin. Esimerkiksi kieliopin tai sanaston keruu ei välttämättä onnistu samoilta kielenoppailta, jotka ovat hyviä kertojia ja folkloren taitajia. Eri ikä- ja sukupuoliryhmillä on usein erilainen kompetenssi kielessään, ja heillä saattaa olla myös erilainen

kaksikielisyysprofiili. Äärimmäisinä tapauksina ovat ryhmät, joissa eri sukupuolet puhuvat eri kieliä.

**Osallistuva havainnointi ja elisitointi.** Ihannetapauksessa lingvisti saavuttaa käytännön taidon tutkimassaan kielessä eli pystyy käyttämään sitä suoraan kenttätyön apukielenä. Tässä piilee kuitenkin se vaara, että tutkija antaa oman kompetenssinsa vaikuttaa aineistoon. Tästä syystä apuna käytetään usein tutkijan ja kielenpuhujien yhteistä metakieltä – esimerkiksi Venäjän alueella venäjää, Kiinassa kiinaa ja Latinalaisessa Amerikassa espanjaa tai portugalia. Metakielen käyttämisessäkin on kuitenkin ongelmansa, sillä sitä voidaan käyttää ainoastaan kaksikielisten puhujien kanssa ja sen tuoma interferenssi saattaa myös vaikuttaa kerättyyn kieliaineistoon. Joissakin tapauksissa kielentutkijan on paras pitäytyä mahdollisimman huomaamattomassa havainnoinnissa, jonka apuna voidaan käyttää myös teknisiä laitteita. Yksi mahdollisuus on antaa kielenoppaiden itsensä dokumentoida puhettaan ilman tutkijan välitöntä läsnäoloa.

Viime kädessä kaikessa kielitieteellisessä kenttätyössä on tehtävä valinta passiivisen havainnoinnin ja aktiivisen kyselyn välillä. Niin sanotulla osallistuvalla havainnoinnilla (engl. *participant observation*) lingvisti – kuten antropologikin – kerää empiiristä aineistoa mahdollisimman monenlaisista konteksteista elämällä tutkimansa yhteisön sisällä neutraalina tarkkailijana. Tutkija pyrkii keräämään tutkimastaan kielestä mahdollisimman laajoja ja edustavia tekstejä. Muun muassa sanaston ja kieliopin kokonaisvaltainen dokumentaatio olisi kuitenkin mahdotonta ilman elisitaatiota eli kielenoppaille suunnattua tarkennettua kyselyä, jonka avulla voidaan täydentää muuten puutteellisiksi jääviä tietoja kielen substanssista ja rakenteesta. Vaikka neopositivistinen kielentutkimus piti elisitaation kautta saatuja tietoja ”epäaitoina”, tiedot kielestä jäisivät poikkeuksetta puutteellisiksi ilman tätä menetelmää. Nykyään elisitaatiota voidaan menestyksellisesti käyttää myös internetin välityksellä.

Kieliaineiston lisäksi antropologinen lingvisti kerää mahdollisimman monipuolista tietoa myös kielen kontekstista. Tätäkään tietoa ei välttämättä saa pelkällä havainnoinnilla, vaan apuna täytyy käyttää haastatteluja. Haastattelut voivat koskea esimerkiksi kielen käyttötilanteita tai

yksilön henkilöhistoriaa, jolloin voidaan rakentaa niin sanottuja kielellisiä elämäkertoja (engl. *linguistic biographies*) eli historiallisia kuvauksia siitä, miten ja missä yhteyksissä yksilö ja hänen lähipiiriinsä kuuluvat muut yksilöt ovat elämänsä aikana käyttäneet omaa kieltään ja muita kieliä ja miten tämä tilanne on muuttunut sukupolvesta toiseen.

**Tilastollinen tiedonkeruu**. Yksi antropologisen lingvistiikan keskeisistä tutkimuskohteista on kielen suhde kieliyhteisöön eli kieltä puhuvaan väestöön. Kokonaiskuvan hankkiminen väestötason kielenkäytöstä edellyttää usein laajamittaista kyselyä, joka voidaan toteuttaa vain systemaattisten ryhmähaastattelujen ja kyselykaavakkeiden avulla. Verrattuna yksilöiden yksityiskohtaisiin syvähaastatteluihin laajamittaiset kyselyt voivat antaa vain huomattavasti yleiskatsauksellisemman käsityksen kielestä ja sen käytöstä, mutta sen etuna on menetelmän suurempi kattavuus ja tiedon yleistettävyys.

Laajamittaisen kyselyaineiston käsittely vaatii demografisten ja tilastollisten analyysimenetelmien käyttöä. Erityisesti demografia eli väestöjen (kr. *dêmos*) tutkiminen on antropologisen lingvistiikan olennainen aputiede. Maailmassa tällä hetkellä puhuttujen kieliyhteisöjen demografinen profiili on hyvin vaihteleva: on kieliä, joiden puhujia on satoja miljoonia, ja toisia, joita puhuu vain muutama yksilö. Valtaosalla maailman kielistä on muutamasta sadasta muutamaan tuhanteen puhujaa. Tämä suuri vaihtelu, joka ilmenee myös puhujien kaksi- ja monikielisyyden eroissa, tarjoaa haasteita tilastoaineiston keruulle ja käsittelylle. Mitä suurempi kieliyhteisö on, sitä vaikeampi siitä on luoda edes tilastollista kokonaiskuvaa.

**Aineiston tallentaminen**. Klassinen ja edelleen käyttökelpoinen menetelmä tallentaa kenttätyössä tehtyjä havaintoja on kirjoittaa ne muistiin, mieluiten järjestelmällisellä tavalla ja tiiviissä yhteistyössä kielenoppaan kanssa. Aineiston segmenttirakenteen osoittamiseen käytetään yleensä vakiintuneita foneettisia tarkekirjoitusjärjestelmiä, kohdekielestä riippuen esimerkiksi kansainvälistä foneettista aakkostoa (*International Phonetic Alphabet* eli IPA) tai suomalais-ugrilaista tarkekirjoitusta (*Finno-Ugrian Transcription* eli FUT). Monista sittemmin kadonneista kielimuodoista on jäljellä ainoastaan lingvistien tekemiä kenttämuistiinpanoja, ja suuri määrä tällaisia muistiinpanoja odottaa

edelleen arkistoissa käsittelijäänsä. Kaikki, varsinkaan varhaisemmat kenttätutkijat, eivät kuitenkaan ole käyttäneet yksiselitteistä foneettista notaatiota, minkä takia aineistot ovat jälkikäteen usein vaikeasti tulkittavissa.

Koska kirjallisten muistiinpanojen luotettavuus riippuu aina niiden tekijän tarkkuudesta ja koska varsinkin tekstien muistiinpaneminen vaatii suurta nopeutta, alettiin jo 1800-luvun lopulla käyttää teknisiä äänentallennuslaitteita ja kokeellis-foneettisia apuvälineitä (ks. Kokeellinen fonetiikka tk.). Vaikka nämä tarjoavat sen ilmeisen edun, että aineisto on myös muiden tutkijoiden ulottuvilla mahdollisimman alkuperäisessä muodossa, tallennusten litteroiminen jälkikäteen on hyvin työlästä ja vaatii tutkijalta joko erittäin hyvää kohdekielen taitoa tai uutta yhteistyötä kielenoppaan kanssa. Ongelmana on ollut myös se, että tallennusformaatit muuttuvat nopeasti eikä vielä hiljattain käytetyillä laitteilla tehtyjä tallennuksia voida välttämättä enää avata. Digitoiminen ei näytä tuovan tähän ongelmaan pysyvää ratkaisua, koska myös digitoinnin formaatit vaihtuvat nopeassa tahdissa ja digitoitu aineistokin on altis fyysiselle kulumiselle.

Joka tapauksessa yleinen suositus on nykyään, että kenttälingvistit käyttävät kenttätyössään audiovisuaalisia apuvälineitä. Näin saatu aineisto voidaan edelleen tallentaa avoimiin lingvistisiin tietojärjestelmiin, kuten eurooppalaiseen korpusprojektiin (European Distributed Corpora Project eli EUDICO), jolla puolestaan on oma äänne- ja muotorakenteen merkintää varten kehitetty visualisointijärjestelmänsä (EUDICO Linguistic Annotator eli ELAN). Suomessa kieliaineistoa voi tallentaa FIN-CLARIN-kielipankkiin, joka kuuluu laajempaan kansainväliseen kielitieteiden CLARIN ERIC -tutkimusinfrastruktuuriin (dokumentaatio- ja tallennusmenetelmistä lisää ks. Kielten dokumentointi ja kieliopin kuvaus tk.).

**Lingvistiset menetelmät**. Varsinaisen kieliaineiston käsittelyyn antropologinen lingvistiikka käyttää tavanomaisia kielitieteellisen analyysin menetelmiä. Erotukseksi monien teoreettisten koulukuntien lähestymistavoista antropologinen lingvistiikka pyrkii kuitenkin minimoimaan teoreettisten ennakko-oletusten haitalliset vaikutukset. Kokemus osoittaa, että monet kieliteoriat ovat ohimeneviä muotivirtauksia, joille on

tyypillistä vaikea terminologia, liiallinen formalismi ja dogmatismi. Kun muotivirtaus on ohi, tällaiselta perustalta tehtyä kielen kuvausta on vaikea suhteuttaa yleiseen tietopääomaan. Tästä syystä antropologinen lingvistiikka – kuten myös kielitypologia – kuvaa kieliä mahdollisimman yksinkertaisilla malleilla, jotka lisäksi ottavat huomioon kielten erilaisuuden.

Antropologisen lingvistiikan kuvausmalliksi sopiikin parhaiten niin sanottu peruskielioppiteoria (engl. *Basic Linguistic Theory*), jossa operoidaan vanhoilla tutuilla kielioppitermeillä (esimerkiksi muotoryhmien ja lauseenjäsenten nimet), yleisesti tunnetuilla perusteorioilla (esimerkiksi fonologinen teoria) ja tavanomaisilla analyyttisillä metodeilla (esimerkiksi morfologinen ja syntaktinen analyysi). Tätä teorianeutraalia kieliteoriaa on puolustanut erityisesti australialainen lingvisti R. M. W. Dixon (2009). Lähtökohtana on, että kieliaineiston hallinnan tulisi käydä teorioiden edellä. Tarkoituksena on saattaa aineisto sellaiseen muotoon, että se on sekä eri lingvististen teorioiden kannattajien että tarvittaessa myös ei-lingvistien helposti ymmärrettävissä. Näin analysoitua aineistoa voi tietysti myöhemmin käyttää myös monimutkaisemman teorianmuodostuksen pohjana.

**Filologiset menetelmät**. Vaikka antropologisessa lingvistiikassa aineisto tyypillisesti kerätään kenttätyön avulla suoraan elävien kielten puhujilta, siinä voidaan hyödyntää myös kirjoitettuun muotoon tallennettuja kieliaineistoja, jolloin käytetään filologian menetelmiä. Erotukseksi lingvistiikasta filologia käsitetään usein tieteenalaksi, joka keskittyy kirjoitettujen lähteiden eli tekstien ja tekstikorpusten sisältöön esimerkiksi historialliselta tai kirjalliselta kannalta (ks. Näkökulmia filologiseen tutkimukseen tk.). Kirjallista aineistoa voidaan kuitenkin tutkia myös lingvistisesti, ja silloin liikutaan antropologisen lingvistiikan alueella. Mitä vanhemmasta aineistosta on kysymys, sitä suurempi merkitys lingvistiikalla on aineiston ymmärtämiselle.

On monia muinaisia kieliä, jotka tunnetaan ainoastaan kirjallisista lähteistä. Vain harvoja näistä kielistä – esimerkiksi etruskia – on kirjoitettu nykypäivään asti käytetyillä kirjoitusjärjestelmillä. Tyypillisempää on, että sekä kieli että sen kirjoitusjärjestelmä ovat jääneet pois käytöstä. Tällöin lingvistin ensimmäinen tehtävä on desifroida eli tulkita

tuntematon kirjoitusjärjestelmä. Tällä saralla onkin saavutettu hämmästyttäviä tuloksia: monet tärkeät muinaiset kielet – esimerkiksi muinaisegypti, sumeri, akkadi, heetti, arkaainen kreikka, muinaisturkki ja kitaani – tunnetaan vain sen ansiosta, että niiden kirjoittamiseen käytetyt kirjoitusjärjestelmät on onnistuttu avaamaan. Näin saadulla aineistolla on poikkeuksellisen suuri arvo myös lingvistisesti, sillä se antaa kuvan muinaisesta kielidiversiteetistä ja sen kehityksestä aikojen kuluessa.

**Diakroniset menetelmät**. Antropologisen lingvistin keräämä synkroninen kenttäaineisto toimii yleisenä tietopohjana tutkittavan kielen tuntemukselle, ja tätä aineistoa voidaan käyttää monenlaisiin sekä tieteellisiin että käytännöllisiin tarkoituksiin. Yksi merkittävä tieteellinen käyttöalue on diakroninen eli historiallinen kielentutkimus, jonka avulla lähestytään tutkittavan kielen aikaisempia vaiheita ja kehitystä. Sikäli kuin tukena ei ole käytettävissä filologista aineistoa, tutkimus käyttää niitä historiallis-vertailevan kielentutkimuksen menetelmiä (ks. Historiallinen ja vertaileva kielentutkimus tk.), jotka kehitettiin Saksassa 1800-luvun alkupuolella ja joita siellä aluksi sovellettiin indoeurooppalaisten kielten tutkimukseen. Tutkimusalue on edelleen ajankohtainen, sillä maailmassa on useita satoja kielikuntia ja suurta osaa niistä ei vielä ole edes alustavasti analysoitu.

Voi sanoa, että historiallinen ja vertaileva kielentutkimus on vain yksi laajasti käsitetyn antropologisen lingvistiikan osa-alueista. Siinä missä kenttätyö antaa kielestä synkronisen kuvan, historiallinen ja vertaileva tutkimus avaa ikkunan diakroniaan. Jos kielellä on sukukieliä, näiden keskinäisellä vertailulla voidaan rakentaa käsitys kielen ja sen sukukielten aikaisemmista vaiheista ja rekonstruoida eritasoisia kantakieliä. Jos kielellä ei ole säilyneitä sukulaisia eli jos se on genealoginen isolaatti, sen sisältämä synkroninen informaatio – esimerkiksi morfofonologia – voi silti antaa mahdollisuuden diakroniseen analyysiin sisäisen rekonstruktion menetelmällä. Myös vertailu erisukuisiin naapurikieliin, joiden kanssa kieli on ollut historiallisessa kontaktissa, antaa diakronista tietoa kielen kehityksestä ja kronologiasta.

# 4. Antropologisen lingvistiikan teoriapohja

Antropologinen lingvistiikka on perusluonteeltaan empiirinen tieteenala, jonka ei päämääräänsä toteuttaakseen tarvitse välttämättä mennä aineiston dokumentaatiota, järjestelyä ja tallentamista pitemmälle. Kielitieteen sisällä kulkee vanha juopa, jonka toisella puolella ovat ne kielitieteilijät, jotka tutkivat ”kieltä” (engl. *language*), ja toisella ne, jotka tutkivat ”kieliä” (*languages*). Antropologinen lingvistiikka tutkii luonnollisesti kieliä ja niiden kautta kielten moninaisuutta ja kohdistaa huomionsa erityisesti niihin piirteisiin, jotka erottavat kieliä toisistaan, eikä niihin, jotka universaalisti yhdistävät kaikkia kieliä.

Mitä enemmän maailman kieliä on tutkittu, sitä ilmeisemmäksi on käynyt, että kielten välillä on hämmästyttävän suuria eroja. Vastaavasti niin sanottujen kielellisten universaalien eli kaikille kielille yhteisten ominaisuuksien määrä on osoittautunut hyvin pieneksi, minkä lisäksi yhteiset piirteet ovat tyypillisesti hyvin triviaaleja ja koskevat lähinnä vain kaikkein perustavimpia kielellisiä kategorioita (esimerkiksi vokaalit vs. konsonantit, äännesegmentit vs. sanat vs. fraasit). Toisaalta myös hyvinkin keskeiset kielen osa-alueet (esimerkiksi tavurakenne, sanajärjestys, sanaluokat, muoto-oppi, lauserakenne) voivat maailman kielissä olla hyvin erilaisia (näitä eroja tutkii kielitypologia, ks. Kielitypologian menetelmät tk.). Tämä erilaisuus merkitsee myös sitä, että ne kieliteoriat, jotka perustuvat vain yhden kielen – esimerkiksi englannin kielen – faktoihin, eivät voi olla yleispäteviä.

Koska kuitenkin antropologisen lingvistiikan kohteena ei ole ainoastaan kieli vaan myös sen konteksti, tämän tieteenalan omat teoreettiset kysymyksenasettelut koskevat nimenomaan kielen ja sen kontekstin välistä suhdetta. Konteksti puolestaan sisältää sekä luonnonympäristön että kulttuurin ja sosiaaliset suhteet. Viime kädessä kontekstiin kuuluu myös ajattelu.

**Kielellinen relativismi**. Kaikkein keskeisin teoreettinen kysymyksenasettelu antropologisen lingvistiikan historiassa koskeekin kielen ja ajattelun suhdetta. Toisin kuin joissain kielentutkimuksen suuntauksissa, joissa lähdetään siitä, että ihmislajin ajattelu toimii universaalisti

samoista lähtökohdista, antropologinen lingvistiikka pitää mahdollisena sitä, että kieli vaikuttaa ajatteluun, toisin sanoen, että ihmiset ajattelevat eri tavoin riippuen siitä, mitä kieltä he puhuvat. Tämä niin sanotun kielellisen relativismin näkökulma tunnetaan myös nimellä Sapir–Whorf-hypoteesi (engl. *Sapir–Whorf hypothesis*) sitä ensimmäisinä kehittäneiden amerikkalaisten lingvistien Edward Sapirin (1884–1939) ja Benjamin Whorfin (1898–1941) mukaan (Sapir 1949; Whorf 1956).

Kielellinen relativismi näkyy käytännössä monissa kielenkäytön yksityiskohdissa. Esimerkiksi viittausjärjestelmät, joilla kielen puhujat viittaavat aikaan ja paikkaan käyttäen deiktisiä sanoja (erityisesti pronomineja) tai muotoja, vaihtelevat suuresti kielestä toiseen. Puhujan, joka hallitsee vain oman kielensä viittausjärjestelmän, on usein vaikea hahmottaa toisen kielen viittausjärjestelmän sisäistä logiikkaa. Klassisen esimerkin saman fyysisen todellisuuden erilaisista kielisidonnaisista hahmotusmalleista tarjoavat myös väritermit, joiden määrä ja suhde värikarttaan vaihtelee. On lisäksi todettu, että värijärjestelmä täydentyy eri kielissä uusilla väritermeillä tietyn, lähes universaalin kaavan mukaan. Tyypillinen on esimerkiksi Itä-Aasian kielissä vallitseva viiden väritermin järjestelmä (valkoinen – musta – punainen – keltainen – sininen/vihreä), kuten Brent Berlin ja Paul Kay (1969) ovat klassisessa tutkimuksessaan osoittaneet.

Johtopäätöksiä kielen vaikutuksesta ajatteluun ei kuitenkaan pidä viedä liian pitkälle. Joskus on esitetty, että jopa kielen perustypologia, esimerkiksi sanajärjestys, voi vaikuttaa yksilön ja yhteisön käyttäytymiseen tai vaikkapa onnettomuusalttiuteen eri tilanteissa. Näin ei varmaankaan ole. Yhtä epätodennäköistä on, että jos kielessä ei ole tempusjärjestelmää, sen puhujilta puuttuisi käsitys ajasta. Kysymys on enemmänkin siitä, että eri kielet ilmaisevat eri asioita eri tavoin, ja asiat, jotka yhdessä kielessä ilmaistaan kieliopin keinoin, voidaan toisessa kielessä ilmaista sanaston avulla tai jättää kokonaan ilmaisematta. Se, mitä kategorioita kieli ilmaisee, ei välttämättä korreloi sen kanssa, kuinka kielen puhujat hahmottavat todellisuuden.

**Ympäristön heijastuminen kieleen**. Vaikka on siis epätodennäköistä, että kieli suoranaisesti vaikuttaisi ajatteluun, on selvää, että ympäristö vaikuttaa kieleen, ja tämän vaikutuksen analysointi

kuuluu antropologisen lingvistiikan keskeisiin teoreettisiin teemoihin. Jo Ferdinand de Saussure (1857–1913), jota tosin ei varsinaisesti voi pitää antropologisen lingvistiikan edustajana, esitti, että ”kansakunnan tavat heijastuvat sen kieleen ja kieli muokkaa kansakuntaa” (de Saussure 2014). Yksiselitteisin yhteys vallitsee ympäristön ja sanaston välillä: jokaisessa kielessä on runsaasti erikoistunutta sanastoa, joka vastaa sen puhujien luonnonympäristöä ja kulttuuri-identiteettiä. Tunnettua on, että esimerkiksi eskimokielissä on monipuolinen lumisanasto. Metsästykseen tai kalastukseen erikoistuneilla kieliyhteisöillä on vastaavasti näihin elinkeinoihin liittyvää erikoissanastoa, jota on mahdotonta kääntää suoraan muunlaista elintapaa edustavien väestöjen puhumille kielille.

Sanaston sisällä erottuu myös suljettuja järjestelmiä, jotka ovat erityisen kulttuurisidonnaisia. Paljon tutkittu antropologien ja siis myös lingvistien kohde on sukulaisuussanasto (engl. *kinship terminology*), joka heijastaa kieliyhteisön sosiaalisia suhteita (Morgan 1871). Sosiaaliset erot ja hierarkiat voivat näkyä myös siinä, miten sanoja käytetään ja miten eri yhteisön jäseniä ja ryhmiä puhutellaan missäkin tilanteessa: esimerkiksi antamista ja saamista tarkoittavat sanat voivat vaihdella sen mukaan, mitkä antajan ja saajan sosiaaliset suhteet ovat. Monissa yhteisöissä on myös eroja naisten ja miesten kielten välillä. Sosiaalisia suhteita osoitetaan kielellisin keinon muun muassa monissa Itä-Aasian kielissä (esimerkiksi japanissa ja koreassa). Eroja naisten ja miesten kielessä on niin ikään joissakin Itä- ja Pohjois-Aasian kielissä (esimerkiksi japanissa ja tshuktshissa).

Lingvistisesti erityisen mielenkiintoisia ympäristön kielivaikutukset ovat silloin, kun niillä on yhteys kielen rakenteeseen. Sosiaaliset suhteet voivat ilmetä sanaston lisäksi myös kieliopissa erilaisina kohteliaisuustasoina, joilla voi olla esimerkiksi morfologinen ilmiasu (Brown & Levinson 1987). Tavallista on myös, että kielissä on genus- ja luokkajärjestelmiä, joiden avulla sanasto luokitellaan fyysisen sukupuolen tai muun ulkoisen ominaisuuden mukaan luokkiin. Luokilla voi olla morfologiset indikaattorit, ja ne voivat kongruenssin kautta ilmetä myös syntaksissa. Lisäksi eri luokkiin viittaaminen saattaa tapahtua erityisen klassifikaattorijärjestelmän avulla, ja myös deiktiset sanat, kuten pronominit, korreloivat yleensä luokkien kanssa (Aikhenvald & Mihas 2019).

Paljon kiistelty kysymys on, voiko myös luonnonympäristö vaikuttaa kielen rakenteeseen. On esitetty, että esimerkiksi kylmässä ilmanalassa puhuttujen kielten äännejärjestelmät välttävät huuliäänteitä (labiaaleja), kuten näyttää olevan tilanne eskimo- ja aleuttikielissä. On myös ehdotettu, että agglutinatiivinen kielityyppi olisi erityisen tavallinen nomadikansojen kielissä, jolloin se korreloisi näiden kansojen tyypillisesti aukeiden luonnonympäristöjen sekä kieliyhteisöjen laajojen nautinta-alueiden kanssa. Toistaiseksi evidenssi tällaisten korrelaatioiden puolesta näyttää olevan heikkoa.

**Kieli ja etnos ajassa ja paikassa**. Kieli on laajalti hyväksytyn näkemyksen mukaan tärkein etnisyyden tunnusmerkki. Diakronisen kielentutkimuksen menetelmät antavat mahdollisuuden seurata kieltä parhaassa tapauksessa useita tuhansia vuosia ajassa taaksepäin. Etnisyyttä määrittävät kuitenkin myös kieliyhteisön kulttuuriset piirteet ja biologinen perimä. On hyvä ymmärtää, että kieli, kulttuuri ja biologiset ominaisuudet eivät välttämättä periydy samassa linjassa vaan enimmäkseen toisistaan riippumatta. Tästä syystä kieliyhteisön etnistä identiteettiä ei voida seurata taaksepäin yhtenäisenä kokonaisuutena, vaan kielellä, kulttuurilla ja biologisilla ominaisuuksilla on kullakin oma taustansa. Lisäksi vain kielen taustaa voidaan seurata yhtenäisenä linjana, jonka määrittää sen polveutuminen aikaisemmista kielimuodoista ja kantakielistä.

Välittömästi historiallis-vertailevan kielentutkimuksen synnyttyä heräsi myös pyrkimys yhdistää kielten ja kielikuntien varhaisvaiheita konkreettisiin aikoihin ja paikkoihin sekä myös muinaisiin väestöihin. Tämä kysymyksenasettelu on edelleen relevantti, ja sitä voidaan pitää antropologisen lingvistiikan tärkeimpänä diakronisena teemana. Paitsi yksittäisten kielten ja kielikuntien historiaa tutkimus pyrkii selvittämään myös kokonaisten maantieteellisten alueiden kielellistä taustaa. Vaikka kaikilla kielillä on puhuma-alueensa, nämä alueet ovat yleensä sekundaareja ja niiden takana on monimutkainen verkosto kielten liikkeitä ajassa ja paikassa. Kielialueet ovat historian ja esihistorian kuluessa voineet kasvaa (ekspansio), supistua (kontraktio) tai siirtyä (translokaatio), ja kielihistoriallinen tutkimus voi antaa tietoa näistä muutoksista.

Keskeinen ja myös suurta yleisöä kiinnostava kysymys kaikkien kielikuntien kohdalla koskee alkukotia (sa. *Urheimat*) eli sitä muinaista

kontekstia, josta kielikunta on lähtöisin. Tätä kysymystä lähestytään usein kielikunnan sisäisen rakenteen pohjalta, jolloin oletetaan, että kielikunta on syntynyt siellä, missä se ensimmäisen kerran jakautui eli missä sen syvimmät sisäiset erot sijaitsevat. Apua antaa myös paleolingvistiikkana tunnettu tutkimusala, joka erityisesti kantakielen kulttuuri- ja luontosanastoa analysoimalla tekee johtopäätöksiä sitä vastanneen muinaisen kieliyhteisön kulttuurista ja luonnonympäristöstä. Lisäaineistoa saadaan kielikontaktien ja kielellisten substraattien tutkimuksella, joka paljastaa kielten muinaisia liikkeitä sekä nykyisten kielten alueilla aikaisemmin puhuttujen kielten jälkiä.

Parhaassa tapauksessa kielen varhaisvaiheet voidaan sitoa absoluuttisesti ajoitettaviin arkeologisiin konteksteihin, mutta yleensä kielen ja arkeologian suhde on diffuusi eikä suoria rinnastuksia voida välttämättä koskaan tehdä. Kantakielten ajoituksesta antaakin parhaan osviitan rekonstruoitavan sanaston määrä ja luonne. Mitä vähemmän yhteistä sanastoa on, sitä kauempana ajassa kantakieli todennäköisesti on. Ongelmana on kuitenkin se, että sanasto uudistuu vaihtelevalla nopeudella, eikä tämän vaihtelun yleisiä lainalaisuuksia vielä täysin ymmärretä.

## 5. Antropologisen lingvistiikan sovellukset

Antropologisella lingvistiikalla on monia käytännön sovelluksia, jotka liittyvät erityisesti kielten globaalin uhanalaisuuden ilmiöön. Vaikka kieliä on aina sammunut, maailman kielidiversiteetti ei ole koskaan ollut niin uhattu kuin nykyään. Arviot ilmiön laajuudesta vaihtelevat, mutta varovaistenkin ennustusten mukaan yli puolet ja pessimistisemmän näkemyksen mukaan (Krauss 1992) jopa 90 prosenttia maailmassa vielä tällä hetkellä puhutuista kielistä sammuu kuluvan vuosisadan aikana. Tähän uhkaan kielitiede on havahtunut vasta 1980-luvulta alkaen. Tyypillistä on, että suuri yleisö ja poliittiset päättäjät suhtautuvat kielten katoon edelleen välinpitämättömästi, samalla kun luonnondiversiteetin katoaminen on jo tiedostettu ongelmaksi – siitä huolimatta, että se

koskee vain muutamaa prosenttia elävän luonnon moninaisuudesta.

Tilanne on siis hälyttävä ja vaatii nopeita toimia antropologisen lingvistiikan harjoittajilta. Katoavia kieliä on tutkittava, mutta myös niitä puhuvia kieliyhteisöjä on tuettava, ainakin silloin kun yhteisöt itse kokevat kielensä katoamisen ongelmaksi. Ilmiöllä on vahva poliittinen ulottuvuus, sillä ilman valtakoneistojen tukea kielten katoamista on mahdoton pysäyttää. Ongelmana on kuitenkin se, että valtakoneistot – muun muassa kansallisvaltiot ja niiden yhteistyöjärjestöt – ovat perinteisesti juuri ne tahot, jotka ovat edesauttaneet kielten katoamista.

**Uhanalaisuuden tiedostaminen**. Ensimmäinen askel kielten katoamisen hillitsemiseksi on ilmiön kartoittaminen ja tietoisuuden levittäminen sen seurauksista. 1990-luvulta alkaen onkin luotu järjestelmiä ja julkaisukanavia, joihin tietoja uhanalaisista kielistä on voitu syöttää ja joista suuri yleisö voi saada niistä tietoa. Tärkein globaali yritys tällä alalla on ollut Unescon julkaisema uhanalaisten kielten kartasto *Atlas of the World's Languages in Danger*, josta on ilmestynyt sekä painettuja että verkkoversioita. Tämän hankkeen rinnalle on tullut aluksi Googlen tuella toteutettu uhanalaisten kielten luettelo (*Endangered Languages Catalogue* eli ElCat), joka pyrkii toimimaan interaktiivisena verkkosivustona.

Yksittäisten kielten uhanalaisuuden astetta on yritetty arvioida monella eri järjestelmällä, joista yksi on *The Ethnologue* -tietokannan käyttämä EGIDS (*Expanded Graded Intergenerational Disruption Scale*). Karkeasti maailman kielet voidaan jakaa elinvoimaisiin (engl. *viable*), uhanalaisiin (engl. *endangered*), katoaviin (engl. *moribund*) ja kuolleisiin (engl. *extinct*), mutta uhanalaisten kielten tila vaihtelee paljon riippuen muun muassa kieliyhteisön absoluuttisesta koosta, ikärakenteesta, stabiliteetista, liikkuvuudesta, kaksikielisyydestä ja virallisesta statuksesta. Yleensä kielten katoamisen takana on kieliyhteisön syrjäytyminen kielten välisessä kilpailussa. On kuitenkin myös entisiä valtakieliä, jotka ovat menettäneet elinvoimansa, kuten Irlannin virallinen kieli iiri.

Usein kysytään, onko uhanalaisten kielten säilyttäminen tarpeellista ja miksi. Vastaus on sama kuin luonnon moninaisuuden osalta: diversiteetti on itseisarvo, joka voi osoittautua ennakoimattomalla tavalla hyödylliseksi. Joka tapauksessa on varmaa, että niin sanottujen luonnonkansojen kulttuurit sisältävät kieleen sidottua tietoa esimerkiksi kasvi- ja

eläinlajeista, joilla on muun muassa lääketieteellistä käyttöä. Kun kieli katoaa, katoaa myös sen kannattama kulttuuri ja toisin päin, joten uhanalaisten kielten suojelu merkitsee myös kulttuuridiversiteetin suojelua.

**Uhanalaisten kielten dokumentointi**. Vähin, minkä antropologinen lingvistiikka voi kielellisen diversiteetin säilyttämiseksi tehdä, on dokumentoida katoavat kielet mahdollisimman kattavasti. On usein esitetty, että lingvistin – eikä vain antropologisen lingvistin – koulutukseen tulisi aina sisältyä kenttätyöhön perustuva aikaisemmin tutkimattoman kielen dokumentoiminen. Näin esimerkiksi lingvistisen väitöskirjatyön kohteeksi tulisi *aina* valita jonkin uhanalaisen kielen kartoitus. Tämän hetken tilanne on pahaenteinen, sillä kieliä katoaa nopeammin kuin niitä ehditään tutkia. Yhtenä syynä on toisen maailmansodan jälkeen alkanut yleisen kielitieteen yksipuolistuminen, joka on vienyt tutkimuksen pois kielidiversiteetistä kohti aivan muita kysymyksenasetteluja. Tilanne on tosin aivan viime aikoina alkanut tältä osin kehittyä parempaan suuntaan. Suomessakin on tehty jo useita väitöskirjoja, joiden kohteena on ollut aikaisemmin dokumentoimaton uhanalainen kieli.

Uhanalaisten kielten kartoitus on mahdollista vain pitkäjänteisen ja systemaattisen kenttätyön avulla, ja kartoituksen tulee tapahtua – kuten aina antropologisessa lingvistiikassa – mahdollisimman yksinkertaisella ja teorioista riippumattomalla tavalla. Näin saatu aineisto palvelee hyvin myös esimerkiksi historiallis-vertailevan kielentutkimuksen ja kielitypologian tarpeita. Ihannetapauksessa uhanalaisen kieliyhteisön kokonaisvaltaiseen tutkimukseen osallistuu tutkijoita myös muilta tieteenaloilta, kuten fyysisestä antropologiasta, folkloristiikasta, uskontotieteestä, historiasta, arkeologiasta ja etnobotaniikasta. Maailmassa on vain harvoja kieliyhteisöjä, jotka on riittävän monipuolisesti kartoitettu.

**Kielten revitalisointi**. Antropologinen lingvistiikka voi myös konkreettisesti auttaa kieliyhteisöjä ylläpitämään ja elvyttämään kieliään. Kielten elvytys eli revitalisaatio on varsinaisesti vasta 1990-luvulta alkaen noussut yhdeksi kielitieteen sovellusalueeksi (Hinton, Huss & Roche 2018). Kielen kato tapahtuu yleensä käyttöyhteyksien eli domeenien menetyksen kautta. Kielen elvytys tulee ajankohtaiseksi viimeistään silloin, kun kielen jatkuvuus yhteisössä on katkennut eli kun on ehtinyt kasvaa yksi sukupolvi, joka ei enää hallitse yhteisön omaa alkuperäistä kieltä. Koska

varsinkin pienten niin sanottujen alkuperäiskansojen kielten käyttöyhteydet ovat hyvin kulttuurisidonnaisia, kielen elvytys vaatii usein myös kulttuurin elvytystä tai ainakin suojelua.

Onnistuneita kielenelvytysprojekteja on maailmassa toistaiseksi vain muutamia: näitä myönteisiä esimerkkejä ovat Uuden-Seelannin maori, Havaijin saarten havaiji sekä Suomen Lapin inarinsaame, joille kaikille on rakennettu uusi puhujasukupolvi. Inarinsaamen elvytystyötä on käsitellyt väitöskirjassaan Annika Pasanen (2015). Päämenetelminä tässä työssä on käytetty toisaalta niin sanottuja opettaja-oppilasryhmiä (engl. *master-apprentice*), toisaalta kielipesiä (engl. *language nest*). Opettaja-oppilasryhmissä vanhemman, kieltä hyvin taitavan sukupolven edustajat kehittävät vapaan keskustelun avulla nuoremman, kieltä heikommin taitavan sukupolven kielitaitoa. Kielipesissä taas kieltä täysin taitamaton nuorin sukupolvi osallistuu päiväkotimuotoiseen kielikylpyyn, jossa käytetään yksinomaan elvytettävää kieltä.

Kielenelvytys on mahdollista ainoastaan, jos yhteisöllä on riittävä motivaatio kielen palauttamiseen. Usein riittää, että yhteisössä on muutama aktiivinen yksilö, joiden johdolla projekti voi alkaa. Yleensä tarvitaan kuitenkin myös julkista tukea, sekä rahallista (esimerkiksi kielipesien toiminnan aloittamiseen) että moraalista. Julkisella tuella valtayhteisö, jonka käsissä myös valtakoneistot ovat, voi osoittaa arvostavansa elvytettävää kieltä. Kokemus osoittaa, että kielen elvytys voi toimia ensiaskeleena myös kulttuurin elvytykselle. Yhteisön omanarvontunto kohoaa oman kielen taidon kasvaessa, mikä vähentää sosiaalista syrjäytymistä ja saattaa viime kädessä olla valtayhteisölle taloudellisestikin edullista.

**Kirjakielten luominen**. Usein kielenelvytysprojekteihin liittyy myös kielen kirjoitetun käytön edistäminen. Useimmat maailman uhanalaisista kielistä ovat vailla kirjoitusperinnettä, mikä tarkoittaa sitä, että niiden puhujat käyttävät kirja- ja sivistyskielinään muita kieliä. Kirjoitetun kielen periaatteiden luominen sopii hyvin lingvistisen antropologian kohteeksi, sillä se edellyttää tutkittavan kielen äännejärjestelmän analyysiä ja sen suhteuttamista käytettävissä oleviin kirjoitusjärjestelmiin, jotka eivät suinkaan aina rajoitu latinalaiseen kirjaimistoon. Tuloksena on parhaassa tapauksessa uuden, kirjallisen domeenin syntyminen uhanalaiselle kielelle.

Olisi kuitenkin virhe kuvitella, että kirjoitetun kielen luominen sinänsä voisi pelastaa kielen katoamiselta. Erityisen tarpeetonta ja jopa vaarallista on luoda sisäisesti aina varioivalle ja aikaisemmin vain puhuttuna käytetylle kielelle yhtenäinen normatiivinen kirjakieli. Sen sijaan että puhujat kokisivat kirjakielen uudeksi domeeniksi, he saattavat nähdä sen omalle puheelleen vieraana ja keinotekoisena rakennelmana. Parempi onkin, jos kielelle luodaan ainoastaan löyhät oikeinkirjoituksen periaatteet, joita voi tarvittaessa soveltaa esimerkiksi sähköisessä kommunikoinnissa. Kirjallisen kulttuurin luominen hyvin pienen yhteisön kielelle on käytännössä epätarkoituksenmukaista ja resursseja haaskaavaa.

**Eettiset kysymykset**. Kielten elvyttäminen ja kirjoitusjärjestelmien luominen voidaan selkeästi nähdä antropologisen lingvistiikan ”lahjoina” tutkittavia kieliä puhuville yhteisöille. Yleensä yhteisöt ottavat nämä lahjat myönteisesti vastaan, ja ne voidaankin nähdä luonnollisina vastapalveluksina yhteisöille, jotka toimivat tutkimuksen kohteina. On kuitenkin vaikeampi ottaa kantaa siihen, onko myös pelkkä kielten dokumentoiminen lahja puhujayhteisöille. Useinhan kielen dokumentoiminen on tutkijalle meritoivaa toimintaa, jonka kautta voi aueta esimerkiksi tieteellinen ura. Puhujille tutkijan apuna toimiminen voi kuitenkin merkitä vain ylimääräistä vaivaa, josta heille ei välttämättä koidu mitään välitöntä korvausta.

Varsinkin pohjoisamerikkalaisen poliittisen korrektiuden diskurssissa on pienten etnisten ryhmien keskuuteen levitetty ajatusta, että kieli, kuten myös sitä puhuvan yhteisön kulttuuri ja geeniperimä, ovat yhteisön omaisuutta, jota ei tule luovuttaa ulkopuoliseen tutkimuskäyttöön. Tällainen asenne on tutkimukselle haitallinen ja tekee antropologisen lingvistiikan harjoittamisen, kuten muunkin antropologisen tutkimuksen, hankalaksi ja poliittisesti arkaluontoiseksi. On kuitenkin muistettava, että kieli on immateriaalinen ominaisuus, jota ei sellaisenaan voida ottaa pois kieliyhteisöltä, vaikka se dokumentoitaisiin – päinvastoin kielen dokumentoiminen voi auttaa yhteisöä säilyttämään kielensä. Kaikki kielet ovat ihmiskunnan yhteistä omaisuutta, eikä kieliyhteisöillä ole yksinoikeutta omiin kieliinsä.

On silti hyvä miettiä, miten kielenoppaiden työ on parhaiten korvattavissa, sillä kielen akateeminen tutkimus ei välttämättä riitä korvaukseksi

heidän vaivannäöstään. Parhaassa tapauksessa tutkija tulee siinä määrin tutkimansa kieliyhteisön jäseneksi, että kysymys ei aktuaalistu, mutta usein tutkijan on korrektia huolehtia siitä, että kielenoppaat saavat myös aineellisen korvauksen. Tämä on aina syytä ottaa huomioon dokumentaatioprojektien budjettia suunniteltaessa (kielten dokumentaatioon liittyvistä eettisistä ja käytännöllisistä ongelmista ks. Kielten dokumentointi ja kieliopin kuvaus tk.).

## 6. Yhteenveto: antropologisen lingvistiikan toimijoita

Vaikka alalla on pitkät perinteet, lingvistinen antropologia on organisoituna kansainvälisenä tieteenalana varsin nuori. Alan modernit oppi- ja käsikirjat ovat enimmäkseen englanninkielisiä, tärkeimpinä William Foleyn *Anthropological Linguistics* (1997) ja Alessandro Durantin *Linguistic Anthropology* (1997). Esimerkki alan viimeaikaisista sovelluksista etnolingvistiikan nimellä on James W. Underhillin *Ethnolinguistic and Cultural Concepts* (2012). Uhanalaisia kieliä käsittelevät esimerkiksi teokset Austin & Sallabank (2012) ja Rehg & Campbell (2018).

Alan tunnetuin yhdysjärjestö on pohjoisamerikkalainen lingvistisen antropologian seura, Society for Linguistic Anthropology (SAL), joka kuuluu osana Amerikan antropologiseen liittoon, American Anthropological Association. Seura on vuodesta 1959 alkaen julkaissut alan aikakauslehteä nimeltä *Anthropological Linguistics*. Toinen merkittävä toimija on niin ikään pohjoisamerikkalainen kielen esihistorian tutkimuksen seura, Association for the Study of Language in Prehistory (ASLIP), joka vuodesta 1995 alkaen on julkaissut aikakauslehteä *Mother Tongue*. Kenttätyöperinteen kehittäjänä on muistettava myös Summer Institute of Linguistics (SIL), joka järjestää säännöllisesti kielitieteen koulutusta protestanttista lähetystyötä tekeville lingvisteille sekä julkaisee kenttätyöhön perustuvia tutkimustuloksia muun muassa sivustolla nimeltä SIL Forum for Language Fieldwork.

Alan suomalaista perinnettä edustava Suomalais-Ugrilainen Seura on perustamisestaan (1883) lähtien julkaissut kielitieteellisen kenttätyön tuloksia. Seuran aikakauskirjassa *Journal de la Société Finno-Ougrienne* ja sen rinnalla ilmestyvissä ”Toimituksissa”, *Mémoires de la Société Finno-Ougrienne*, on jo 135 vuoden ajan ilmestynyt lukuisien euraasialaisten kielten kieliopillisia ja kielihistoriallisia kuvauksia, joita täydentävät seuran julkaisemat tieteelliset sanakirjat. Seura on myös merkittävällä tavalla osallistunut kielten uhanalaisuutta koskevaan keskusteluun jo paljon ennen kuin aihe nousi kansainvälisen lingvistiikan kohteeksi.

**Aiheesta lisää:**

Duranti, Alessandro. 1997. *Linguistic Anthropology*. Cambridge: Cambridge University Press.

Foley, William. 1997. *Anthropological Linguistics: An Introduction*. Malden (MA): Blackwell.

Underhill, James W. 2012. *Ethnolinguistics and Cultural Concepts: Truth, Love, Hate and War*. Cambridge: Cambridge University Press.

## KIRJALLISUUS


Aikhenvald, Aleksandra Y. & Mihas, Elena (toim.). 2019. *Genders and Classifiers: A Cross-Linguistic Typology.* Oxford: Oxford University Press.

Austin, Peter & Sallabank, Julia (toim.). 2012. *The Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press.

Boas, Franz. 1911–1922. *Handbook of American Indian languages*. Smithsonian Institution, Bureau of American Ethnology Bulletin 40. Washington: Government Printing Office.

Berlin, Brent & Kay, Paul. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.

Bopp, Franz. 1816. *Über das Conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprache*. Frankfurt am Main: In der Andreäischen Buchhandlung.

Brown, Penelope & Levinson, Stephen. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.

Dixon, R. M. W. 2009. *Basic Linguistic Theory*, vol. 1–2. Oxford: Oxford University Press.
Duranti, Alessandro. 1997. *Linguistic Anthropology*. New York (NY): Cambridge University Press.
Foley, William A. 1997. *Anthropological Linguistics: An Introduction*. Oxford: Wiley Blackwell.
Hinton, Leanne, Huss, Leena & Roche, Gerald (toim.). 2018. *The Routledge Handbook of Language Revitalization*. New York (NY) & London: Routledge.
Krauss, Michael. 1992. The world's languages in crisis. *Language* 68:1, 4–10.
Morgan, Lewis H. 1871. *Systems of Consanguinity and Affinity of the Human Family*. Contributions to Knowledge, vol. XVII. Washington: Smithsonian Institution.
Mühlhäusler, Peter. 1996. *Linguistic Ecology: Language Change and Linguistic Imperialism in the Pacific Region*. London & New York (NY): Routledge.
Pasanen, Annika. 2015. *Kuávsui já peeivičuovâ. 'Sarastus ja päivänvalo'. Inarinsaamen kielen revitalisaatio*. Uralica Helsingiensia 9. Helsinki: Helsingin yliopisto & Suomalais-Ugrilainen Seura.
Rehg, Kenneth L. & Campbell, Lyle (toim.). 2018. *The Oxford Handbook of Endangered Languages*. Oxford: Oxford University Press.
Sapir, Edward. 1949. *Selected Writings in Language, Culture and Personality*. Toimittanut David Mandelbaum. Berkeley: University of California Press.
Saussure, Ferdinand de. 2014. *Yleisen kielitieteen kurssi*. [*Cours de linguistique générale*.] Tampere: Vastapaino.
Underhill, James W. 2012. *Ethnolinguistics and Cultural Concepts: Truth, Love, Hate and War*. Cambridge: Cambridge University Press.
Whorf, Benjamin. L. 1956. *Language, Thought, and Reality: Selected Writings of Benajmin Lee Whorf*. Toimittanut John B. Carroll. Cambridge (MA): MIT Press.

## ERÄITÄ VERKKO-OSOITTEITA

Association for the Study of Language in Prehistory: http://aslip.org/MT/MT1.htm.
Endangered Languages Catalogue: http://ling.hawaii.edu/research-current/projects/elcat/.
SIL International: https://www.sil.org.
Society for Linguistic Anthropology: http://linguisticanthropology.org.
Suomalais-Ugrilainen Seura: https://www.sgr.fi.
Unescon Atlas of the World's Languages in Danger: http://www.unesco.org/languages-atlas/.

# Osa IV

# Esimerkkitapauksia menetelmällisestä kirjosta kielentutkimuksen alueilla

# Luku 17 Kieliopin tutkimus ja kielioppiteoriat

*Urpo Nikanne*
https://orcid.org/0000-0002-3267-6705
*Geda Paulsen*
https://orcid.org/0000-0002-2749-4528

**Mikä?**

Kieliopin tutkimus on kielen rakenteen kuvausta. Kuten muukin tieteellinen tutkimus kieliopin tutkimus perustuu järjestelmälliseen havainnointiin ja johdonmukaiseen päättelyyn. Tutkimuksen tavoitteena on kuvata yksittäisten kielten järjestelmiä sekä kielten välisiä yhtäläisyyksiä ja eroja. Kieliopin tutkimus on kielentutkimuksen vanhimpia osa-alueita, ja koulukuntaeroja ja eri traditioita on useita.

**Katso myös:**

Luku 1 Kielitaju kielentutkijan työkaluna

Luku 13 Kielten dokumentointi ja kieliopin kuvaus

Luku 14 Kielitypologian menetelmät

Luku 15 Historiallinen ja vertaileva kielentutkimus

Luku 16 Antropologinen lingvistiikka

# 1. Johdanto

Sanalla *kielioppi* voidaan viitata erilaisiin ilmiöihin. Äidinkielen kielioppi on jokaisen kielenkäyttäjän sisäinen tieto siitä, miten kieli rakentuu, millaiset rakenteet noudattavat kielenkäyttäjän oman kieliyhteisön kielellisiä normeja (ks. Kielitaju kielentutkijan työkaluna tässä kirjassa). Tämä tieto käsittää esimerkiksi (implisiittisesti) sisäistetyt säännöt siitä, miten äidinkielen äänteet vaihtelevat tai sanat taipuvat. Kielenpuhujan sisäinen kielioppi eroaa kuitenkin laajemmasta käsitteestä *kielitieto*, joka kattaa kielen rakenteen lisäksi esimerkiksi puhujan mahdolliset tiedot kielen (ja kielten) historiasta, kehityksestä, variaatiosta, kielen yhteyksistä kulttuuriin ja käyttötilanteisiin jne. (Ks. esim. Kalliokoski, Kumenius, Luukka, Mustaparta, Nissilä & Tuomi 2015.)[1] Voikin sanoa, että kielitiedon ja sisäisen kieliopin avulla kielenpuhuja oppii *kielitietoisuutta* – kielen ilmiöiden ja niiden yhteyksien liittymistä käytäntöön (ks. Hakulinen, Kauppinen, Leiwo, Paunonen, Räikkälä, Saukkonen, Yli-Vakkuri, Östman & Alho 1996).

Kielioppia kuvattaessa kieltä lähestytään järjestelmänä. Termillä *kielioppi* voidaankin viitata myös kieliopin kuvaukseen. Kirjoitetut kieliopit palvelevat erilaisia tarkoituksia. Kieliopin päämäärä voi olla ohjeiden antamista hyväksyttävästä ja oikeanlaisesta kielestä. Tällainen lähestymistapa luonnehtii esimerkiksi kielenhuollon ja -opetuksen tarpeisiin laadittuja kielen rakenteen kuvauksia. Näitä yleensä kirjakielen sääntöjä kuvaavia kielioppeja kutsutaan **preskriptiivisiksi**. Kielentutkijan näkökulmasta kielioppi on yleisemmin ihmiskielen ilmausten rakentumisperiaatteita kuvaava järjestelmä, jota voidaan lähestyä hyvin erilaisista tarkastelukulmista sekä erilaisin työkaluin ottamatta kantaa siihen, onko kuvattava kielen ilmiö suositeltavaa tai luvallista. Tällaiset vakiintuneet puhutun ja kirjoitetun kielen käytänteitä kuvaavat kieliopit ovat **deskriptiivisiä**. Myös deskriptiivinen kielioppi kuvaa normeja, koska kieli on luonteeltaan normatiivinen järjestelmä, joka toimii ihmisten välisen viestinnän välineenä (Itkonen 1978). Tieteellisen kielentutkimuksen alaan kuuluu laaja kirjo erilaisia kielioppeja tai suuntauksia: strukturaalinen, generatiivinen, kognitiivinen, systeemis-funktionaalinen,

konstruktio-, dependenssi- ja sijakielioppi, typologien käyttämä peruskielioppi jne. Tässä artikkelissa keskitymme kielioppiin kielen kuvauksen kannalta ja deskriptiivisestä näkökulmasta.

Kielioppi tarkoittaa lyhyesti sanottuna kielen rakenteen kuvausta. Kieliopin kirjoittaminen perustuu ajatukseen, että kielessä on paljon sellaista säännönmukaisuutta, joka voidaan kuvata yleisellä tasolla samaan tapaan kuin vaikkapa luonnontieteiden kaavat ($e = mc^2$) tai lait (esim. Suomen perustuslain 9 § lause *Suomen kansalaisella ja maassa laillisesti oleskelevalla ulkomaalaisella on vapaus liikkua maassa ja valita asuinpaikkansa*. = 'Jokaiselle Suomen kansalaiselle tai maassa laillisesti oleskelevalle ulkomaalaiselle pätee, että hänellä on vapaus liikkua maassa ja valita asuinpaikkansa'). Samaa säännönmukaisuutta voidaan soveltaa erityistapauksiin. Aikaisemmin kuvaamatonta kieltä lähestyessä onkin tapana kirjoittaa toisaalta sanakirja ja toisaalta kielioppi. Luonnollisen kielen kieliopin kuvauksen pohjana on dokumentointi. Christian Lehmann määrittelee kielen dokumentoinnin representatiivisen aineiston kokoamiseksi, esittämiseksi ja työstämiseksi. Representatiivinen tarkoittaa tässä yhteydessä aineistoa, joka antaa selkeän ja riittävän kuvan siitä, miten ja mihin tarkoituksiin kielen rakennetta käytetään, ja joka on myös muiden tutkijoiden ulottuvilla. Kielen kuvauksen tarkoitus on määritellä tämän aineiston pohjalta säännöt, joiden mukaan kieli toimii. (Lehmann 2001, 83; aiheesta lisää ks. Kielten dokumentointi ja kieliopin kuvaus tk.) Joka tapauksessa kielen kuvaaminen kieliopiksi edellyttää, että kielen rakenteeseen sisältyy olennaisesti säännönmukaisuutta. Muussa tapauksessa dokumentointi kannattaisi tehdä muulla tavoin kuin kielioppisäännöillä, kuten tyytymällä luokitteluihin ja kategorioiden luonnehtimiseen.

Kielioppiin kuuluvat yleisen käsityksen mukaan äänneoppi (fonologia), muoto-oppi (morfologia) ja lauseoppi (syntaksi) sekä näiden keskinäinen vuorovaikutus. Rajat kieliopin eri osa-alueiden välillä eivät ole tarkkoja tai itsestään selviä. Joissain kieliopin kuvauksissa esimerkiksi morfologiaa ja syntaksia käsitellään yhtenä kielijärjestelmän osana. Kieliopilla voidaan erikseen tarkoittaa vain kielen säännönmukaisuuksia. Yksi tällainen kaikille tuttu kieliopillinen säännönmukaisuus on suomen kielessä, niin yleiskielessä kuin murteissakin, esimerkiksi se,

että adjektiiviattribuutti kongruoi pääsanansa kanssa sijassa ja luvussa (*iso**lla** talo**lla*** : *iso**illa** talo**illa*** : *iso**ssa** talo**ssa*** : *iso**issa** talo**issa***). Kongruenssi toteutuu riippumatta siitä, mikä substantiivi ja mikä adjektiivi kulloinkin on kyseessä. Toinen tuttu esimerkki on vaikkapa se, että ensitavua kauempana yhdistelmän CV (C on jokin konsonantti, V on jokin vokaali) edellä on aina tavunraja (*ka.**la*** = CV.**CV**, *kal.**la*** = CVC.**CV**, *kars.**ki*** = CVCC.**CV**, kar.**me**.a = CVC.**CV**.V, *lii.**ma*** = CVV.**CV**).

Käsittelemme ensin lyhyesti kieliopin kirjoituksen historiaa sekä yleisesti että suomen kielen kuvauksen kannalta. Sen jälkeen tarkastelemme kieliopin tutkimuksen keskeisiä osa-alueita, joitain eri kielioppitraditioiden keskinäisiä eroja sekä kieliopin soveltamista.

# 2. Kieliopin kirjoituksen historiaa ja taustaa

Kieliopin kirjoitus on historian kuluessa aina ollut kytköksissä yleiseen yhteiskunnalliseen kehitykseen ja vallitseviin tieteellisiin ajatusmalleihin. Vanhin tunnettu kielioppi on sanskritin kielioppi (noin 400 eKr.), mutta jo sumerilaisten tiedetään analysoineen kieltä (ks. Itkonen 1991; Rasinkangas 2018, 18). Kielioppeja laadittiin myös antiikin Kreikassa. Antiikin aikoina kielitiedettä ei harjoitettu omana tieteenalanaan, eikä silloin ollut vielä kieliopin tutkimusta samassa mielessä kuin se nykyisin ymmärretään. Nykyinen kielenkuvaus kuuluu kuitenkin niihin kielentutkimuksen osa-alueisiin, jotka ovat vahvimmin kytköksissä antiikin perinteeseen (ks. Viljamaa 2007). Jo Sokrates (470–399 eKr.), Platon (427–347 eKr.) ja Aristoteles (384–322 eKr.) pohtivat kielen olemusta ja alkuperää. Monet vanhasta Kreikasta periytyneistä kieltä koskevista kysymyksenasetteluista ovat jatkuvasti ajankohtaisia, kuten kysymys siitä, onko kieli katsottava luonnonlakien määräämäksi ilmiöksi (*physis*) vai sosiaaliseksi konventioksi (*nomos*), tai millainen on kielellisen merkin ja todellisuuden suhde. Platon käsittelee viimeistä kysymystä etymologioiden kautta kielifilosofisessa dialogissaan *Kratylos*, jossa hän pohtii, voiko äänneyhdistelmän määritellä jonkin merkityksen luonnolliseksi

vastineeksi. Aristoteles (1994) puolestaan käyttää ominaisuuksia ja relaatioita käsittelevässä logiikassaan termejä, joita ilman nykyinen lauseenjäsennys ei tule toimeen (*predikaatti*, *subjekti*, *objekti*), vaikka niitä alun perin ei ollut tarkoitettu varsinaisesti kielen kuvaukseen.

Ensimmäisessä kreikan kieliopissa noin vuodelta 100 eKr, *Tekhne grammatikessa*, Dionysios Traakialainen erottaa kahdeksan sanaluokkaa: substantiivit, verbit, partisiipit, artikkelit, pronominit, prepositiot, adverbit sekä konjunktiot; verbien kuvauksesta löytyy jo pääluokka, tapaluokka (modus), persoona, luku, aikamuoto (tempus) ja konjugaatio. Suurin ero nykyaikaisiin kieliopin esityksiin on siinä, että lauserakenteen analyysi puuttuu kokonaan – se ilmaantui pari vuosisataa myöhemmin Apollonius Dyscoluksen kieliopissa, jossa käsitellään verbien, pronominien ja prepositioiden käyttäytymistä sanojen yhdistelmissä (esimerkiksi kongruenssia). Ainut meille antiikista periytynyt latinan lauseopin kuvaus on Priscianuksen vuonna 527 kirjoittama *Institutiones grammaticae*, jossa käsitellään riippuvuus- eli dependenssisuhteita, kuten rektiota eli pääsanan suhdetta täydennykseen. Toinen merkittävä myöhäisantiikin grammaatikko oli Aelius Donatus, jonka *Ars minor* (kysymys-vastausmuotoinen sanaluokkien käsittely) ja *Ars maior* (perusteellinen latinan kielen kuvaus, joka sisälsi myös tyyliohjeet) vaikuttivat eurooppalaisiin kielioppikirjoituksiin 400-luvulta 1200-luvulle asti. (Ks. esim. *Encyclopedia of Linguistics*; *Oxford Handbook of History of Linguistics*; Householder 1981.)

Myöhäisantiikin latinan kieliopit voidaan sisältönsä ja rakenteensa perusteella luokitella neljään ryhmään, luonteeltaan joko pedagogisempiin tai teoreettisempiin: **koulukieliopit**, **sääntökieliopit**, **rakenneanalyysikieliopit** ja **kielioppikommentaarit**. Koulukielioppeja luonnehtii hierarkkinen rakenne lukujen sisällä sekä semantiikan korostuminen. Myös sääntökieliopit ovat luokittelevia ja selittäviä mutta sisältävät runsaammin paradigmoja ja esimerkkejä sekä painottavat kielen muotoaspekteja (esim. sanan sukua tai taivutusluokkaa eli deklinaatiota). Huomattavaa on, että sääntökieliopit ovat tulleet erityisesti alueilta, joissa latina ei ollut äidinkieli. Rakenneanalyysikielioppi (engl. *parsing grammar*) käyttää kysymys-vastausmetodia tekstin sanojen ominaisuuksien analyysiin. Kielioppikommentaarien kirjoitus oli suosittua

varsinkin 600–700-luvulla; sillä tarkoitetaan lähinnä äidinkielisille opiskelijoille suunnattuja seikkaperäisempiä analyyseja jostakin kieliopillisesta aiheesta verrattuna kommentoitaviin kielioppeihin. (Law 1997; Luhtala 2013; Harjunpää 2016.) Näitä myöhäisantiikin kielioppityyppejä käytettiin edelleen mallina keskiajan kielioppien kirjoituksessa. Uutena tyyppinä ilmaantui **runomuotoinen kielioppi**, kuten pedagogiseen käyttöön tarkoitettu Eberhardus Bethunensiksen *Graecismus* (Luhtala 2013).

Keskiajan loppuun saakka antiikin kielioppikirjoitus vaikutti vahvasti eurooppalaisessa kielenkuvauksessa tähdäten lähinnä käytännöllisiin ja normatiivisiin tarkoituksiin. Kun kuvauksen kohteena olivat olleet pääasiassa latina ja kreikka, niin renessanssin aikana, kehittyvän kansallisvaltion idean sekä uskonpuhdistuksen myötä ilmestyivät jo ensimmäiset ”kansankielten” kieliopit – espanjan, italian, ranskan, puolan ja muinaiskirkkoslaavin kuvaukset. Ensimmäinen englannin kielioppi ilmestyi 1586. Vanhin säilynyt suomen kielen kielioppi on Turun akatemian ensimmäisen teologian professorin Eskil Petraeuksen laatima *Linguae Finnicae brevis institutio* (’lyhyt suomen kielioppi’) vuodelta 1649. Uudet kieliopit keskittyivät kehittyvien standardikielten kirjoitusasun ja ääntämisen vastaavuusongelmiin sekä fonologisten ja morfologisten järjestelmien kuvaukseen. Uutta oli silloisten vakiintuneiden sanaluokkien kyseenalaistaminen – ehdotettiin sekä tiiviimpiä että yksityiskohtaisempia luokitteluja (nykyisin sanaluokkien kategorisointi saattaa vaihdella yksittäisen kielen sisälläkin näkökulman ja tarkoituksen mukaan). Eurooppalaisten kielten lisäksi julkaistiin Uuden maailman kielten kielioppeja muun muassa taraskista (engl. *Tarascan*; Meksiko) 1558, ketšuasta (Peru) 1650, nahuatlista (Meksiko) 1571 ja guaranista (Paraguay) 1640. 1600–1700-luvulla lisääntyivät japanin, persian ja kiinan kieliopit. (Ks. *Oxford Handbook of History of Linguistics*.) Sitä, miten eri puolilla maailmaa puhuttuja (usein uhanalaisia) kieliä on eri aikoina kuvattu ja miten niihin kuvauksiin on suhtauduttu, käsitellään tarkemmin tämän teoksen artikkelissa Antropologinen lingvistiikka.

Perinteisen kieliopin kehitys moderniin suuntaan lähti 1800-luvun kielihistoriallisesta ja vertailevasta eli komparatiivisesta kielentutkimuksesta. Vuonna 1786 englantilainen kielentutkija William Jones oli huomannut sanskritissa rakenteellisia yhtäläisyyksiä kreikkaan ja

latinaan mutta myös kelttiin, goottiin sekä muinaispersiaan. Tätä havaintoa seurasi vireä kielten kehityksen ja niiden sukulaisuussuhteiden selvittämisen aikakausi; pohjana käytettiin ensin lähinnä äänneopillisia, myöhemmin myös muoto- ja lauseopillisia piirteitä. Esimerkiksi Franz Bopp vahvisti ajatuksen indoeurooppalaisen kieliperheen olemassaolosta vertaillen eri kielten verbintaivutuksia. Hän osoitti, että kaikilla näillä kielillä on sama kantakieli. Rasmus Rask ja Jacob Grimm puolestaan osoittivat, että kielten muutokset tapahtuvat systemaattisesti. Sen jälkeen alettiin laatia rekonstruktioita oletetuista kantakielistä niiden haaroja, tytärkieliä, vertaamalla. Luonnontieteiden mallia mukauttaen ns. nuorgrammaatikoiden koulukunta yritti selvittää kaavoja, joiden mukaan kielen muutoksia voitaisiin selittää. Myös Suomessa vallitsi yleisenä kielitieteellisenä suuntauksena 1800-luvun lopulta aina 1950-luvulle asti historiallinen kielitiede – tutkijoita kiinnostivat suomen rakenteen ja murteiden lisäksi sen sukukielet, joiden rakenteiden taustalta löydettiin yhteisiä piirteitä (ks. Historiallinen ja vertaileva kielentutkimus tk.). Niin sanotun varhaisnykysuomen kaudella (noin 1820–1879), jolloin suomesta kehittyi täysipainoinen sivistyskieli ja maan virallinen kieli (1863), varsinkin opetuskielioppien tarve suureni. Luonteeltaan ja päämääriltään preskriptiivisiä kielioppeja syntyi lisää, ja niitä kirjoitettiin ruotsiksi sekä suomeksi. Aikakauden kielikäsitykset heijastivat itä- ja länsimurteiden välistä ”taistelua” – kun ensimmäiset kieliopit ja raamatunkäännökset perustuivat länsimurteisiin, nyt alkoi itämurteiden aseman vahvistuminen.

Kun Ferdinand de Saussuren luentojen pohjalta laadittu kirja *Cours de linguistique générale* julkaistiin vuonna 1916, historiallis-vertailevan kielitieteen rinnalle kehittyi synkroninen deskriptiivinen kielitutkimus ja kielijärjestelmän kokonaisuuden korostaminen eli strukturalismi. Strukturalistien huomio kiinnittyi fonologian, morfologian ja syntaksin rakenteen sekä kielen omaksumisen säännönmukaisuuksien kuvaukseen. Alettiin myös keskustella kielen relatiivisista ja universaaleista ominaisuuksista eli siitä, ohjaako kieli puhujan ajattelua vai ovatko kielen ominaisuudet päinvastoin ihmisen evoluution ja aivorakenteen tulos. Debatin avasivat yhdysvaltalaiset kielenantropologit Benjamin Lee Whorf ja Edward Sapir relativistien sekä kielitieteilijä ja kognitivisti

Noam Chomsky universalistien osalta. Nämä aiheet puhuttavat kielentutkijoita vielä nykyisinkin. Myös suomen kieliopin kuvaaminen alkoi kehittyä maailmansotien jälkeen, kun kansainväliset kielitieteelliset suuntaukset (kuten Chomskyn kehittämä generatiivinen transformaatioteoria) alkoivat saada enenevässä määrin tilaa tutkimuksessa ja tieteellinen keskustelu monipuolistui vähitellen. Aarni Penttilän vuonna 1957 valmistunut *Suomen kielioppi* oli erikoinen siinä mielessä, että yleiskielen rakenteen kuvaukseen oli sisällytetty myös siitä poikkeavia ja puhutun kielen rakenteita. Osmo Ikola (1964; 1974) ja Paavo Siro (1964) keskittyivät suomen lauseopin kuvaukseen. Siro kuvasi lauserakenteita logiikan sääntöjen pohjalta, autonomisena formaalisena järjestelmänä, korvaten lauseen toistuvat osiot symboleilla (merkeillä) päämääränään käsitteiden mahdollisimman eksakti kuvaus. 1900-luvun lopun ja 2000-luvun aikana kieliopin tutkimus ja muu kielitiede on edelleen monipuolistunut ja kansainvälistynyt. (Suomen kielitieteen historiasta ks. Karlsson 1997.) Vuonna 2004 ilmestyi tähän asti laajin suomen kieliopin esitys, *Iso suomen kielioppi* eli ISK (Hakulinen, Vilkuna, Korhonen, Koivisto, Heinonen & Alho 2004), josta on vuonna 2008 toimitettu myös verkkoversio (VISK).

## 3. Kuinka kieliopin tutkimusta harjoitetaan

Kielioppeja laaditaan lähtien erilaisista kielikäsityksistä eli erilaisista kielen perusluonnetta koskevista oletuksista. Jotta kieliopin laatiminen olisi mielekästä, on kuitenkin taustalla oltava käsitys, että kielen rakenne on edes jossain määrin strukturoitunutta ja säännönmukaista. Muuten sääntöjen laatimiselle ei olisi järkeviä perusteita. Siitä, kuinka kattavasti kielen rakenne voidaan kuvata sääntöjen avulla, on luonnollisesti erilaisia käsityksiä niin tutkimustraditioiden välillä kuin niiden sisälläkin.

## 3.1. Kieliopin tutkimuksen lähtökohta: kielen rakenne järjestelmänä

Kieliopin kirjoittamisen lähtökohtana on, että kielen rakenne ainakin keskeisiltä osiltaan voidaan esittää järjestelmän muodossa siten, että kielen ilmiöt jakaantuvat yleisempiin kategorioihin (esimerkiksi sanaluokkiin) ja samaan kategoriaan kuuluvat ilmiöt (esim. substantiivit) käyttäytyvät samanlaisessa kontekstissa samalla tavalla. Toisin sanoen kategoriat tunnistetaan siltä pohjalta, miten eri elementit käyttäytyvät muodon ja distribuution osalta. Esimerkiksi suomessa toisen substantiivin attribuuttina olevat substantiivit saavat genetiivisijan, jos mikään poikkeussääntö ei ole voimassa (esim. *Maija**n**/huomispäivä**n**/piru**n** työpaikka*).

Käsitys kielen järjestelmäluonteesta vaihtelee vahvuudeltaan eri koulukunnissa. Esimerkiksi Noam Chomskyn (ks. esim. 1965; 1986; 2000) generatiivisessa kielitieteessä kielijärjestelmän ajatellaan perustuvan ihmisen lajityypillisiin biologisiin valmiuksiin. Toiset tutkijat (esim. Foley & van Valin Jr. 1984) olettavat, että kielten kategoriat ovat samankaltaisia, koska ihmisten viestinnän tarpeet perustaltaan ovat samanlaisia kulttuurista riippumatta. Kielen keskeisten osien esittämiseen järjestelmänä voi suhtautua myös käytännöllisemmin: se on tiivis tapa esittää kielen ominaisuuksia mahdollisimman kattavasti riippumatta siitä, olettaako taustalle jotain sellaista järjestelmäluonteen ennalta määräävää tekijää kuin genetiikka tai sellaista järjestystä luovaa tekijää kuin inhimilliset viestinnän tarpeet.

Kielten perusrakenteen kuvaukseen esimerkiksi kielitypologian tarpeisiin laaditaan kielijärjestelmän kuvauksia, joita kutsutaan myös *referenssikieliopeiksi*. Myös nämä mallit edellyttävät, että kielten järjestelmä on olennaisilta osin kaikille kielille yhteinen ja voidaan laatia kielioppimalli, joka käy tuntemattomienkin kielten dokumentaation välineeksi ja jonka avulla kieliä voidaan vertailla yhteisellä mitalla. Vaikka nämä deskriptiiviset kieliopit on laadittu pelkistetyiksi, niihin pätevät siis samat peruslähtökohdat kuin muihinkin kieliopin malleihin. (Referenssikieliopeista tarkemmin ks. Kielten dokumentointi ja kieliopin kuvaus tk.)

## 3.2. Kieliopin tutkimuksen kiinnostuksen kohteita

Kuten luvussa 2 on kerrottu, kieliopin kirjoittamisella on tuhansien vuosien historia takanaan. Niinpä nykyiset kategoriat perustuvat pitkälti jo aiemmin kehitettyihin malleihin. Lauseopissa peruslauseenjäsenet subjekti ja predikaatti sekä pääsanaluokat substantiivi, verbi, adjektiivi jne. ovat antiikin peruja. Nykyään voidaan lähteä tarkastelemaan kieltä näiden perinteisten kategorioiden pohjalta, koska esimerkiksi sanaluokkien morfosyntaktiset ominaisuudet (miten taivutetaan, missä lauseenjäsenasemissa voivat esiintyä jne.) tunnetaan monilta osin hyvin.

Vaikka kieliopin peruskategoriat ovat varsin vakiintuneet ja niiden pääasiallinen käyttäytyminen tunnetaan, silti tutkimusta tarvitaan. Kieliopin tutkimuksella on, yksinkertaistetusti sanottuna, kolme erilaista tutkimuksen aluetta:

1. yksittäisten kielten rakenteellisten ominaisuuksien selvittäminen
2. eri kielten kielioppien keskinäinen vertailu
3. kielioppiteorian kehittäminen.

Seuraavaksi käsittelemme lyhyesti kutakin näistä kolmesta alueesta.

**Yksittäisten kielten rakenteellisten ominaisuuksien selvittäminen**. Yksittäiskielten, vaikkapa suomen kielen, rakenteessa on aina selvitettävää, esimerkiksi kielen rakenteiden keskinäisissä suhteissa. Lisäksi voidaan aina pyrkiä selvittämään yleisempiä ja kattavampia kuvauksia. Tämä tutkimusalue on kaikkein tyypillisin ja tutuin, kuten myös edellä olevasta historiallisesta katsauksesta voi huomata. Yksittäisten kielten huolellinen kieliopillinen kuvaus toimii pitkälti muiden tutkimusalueiden pohjana.

**Eri kielten kielioppien keskinäinen vertailu**. Eri kielissä samat kategoriat käyttäytyvät hieman eri tavoin, vaikkapa sanajärjestyksen kannalta: missä järjestyksessä ovat subjekti ja predikaatti, missä järjestyksessä erilaiset attribuutit suhteessa pääsanaansa. Vaikka yksittäiset kategoriat eivät olekaan kielikohtaisia, niin kategorioiden joukko on. Suomen kieli ei käytä hyväkseen kaikkia niitä fonologisia distinktiivisiä eli jonkin kielen äännejärjestelmän äänteitä toisistaan erottavia piirteitä kuin esimerkiksi ruotsin kieli: esimerkiksi eroa [soinnillinen] vs. [soinniton] ei

suomen kieli käytä erottelemaan foneemeja, vaikka suomessakin toki on soinnillisia ja soinnittomia äänteitä. Sen sijaan ruotsin kielessä sanat ***pil*** 'nuoli' ja ***bil*** 'auto' ovat minimipareja, jotka perustuvat klusiilien soinnillisuuseroon.

Kielten keskinäisten kieliopillisten erojen ja yhtäläisyyksien vertailu on laajimmillaan typologista tutkimusta, ja sen avulla voidaan tehdä päätelmiä, joita ei yhden kielen perustella voi tehdä. Koska aikojen kuluessa on tehty tutkimusta eri kielten sijajärjestelmistä, niin nykyään on mahdollista ymmärtää vaikkapa suomen kielen sijajärjestelmää paremmin kuin vielä joitain satoja vuosia sitten (vrt. luku 2). Kielten rakenteiden vertailun tärkeä infrastruktuuri on nykyään vapaasti verkossa käytettävissä. Tunnetuimpia typologisia tietokantoja on maailman kielten rakenteiden kartasto WALS (*The World Atlas of Language Structures*; Dryer & Haspelmath 2013). Erilaisia typologisia tietokantoja yhdistää esimerkiksi Typological Database System, kun taas AUTOTYP-tietokanta tarjoaa raakadataa prosessoivan systeemin. Typologisesta kielitieteestä kerrotaan lisää tämän teoksen artikkelissa Kielitypologian menetelmät (ks. myös Kolehmainen, Miestamo & Nordlund 2013).

Kapeamman näkökulman kielten vertailussa ottaa kontrastiivinen kieliopin kirjoitus, jonka tavoitteena on tutkia yleensä kahden kielen välisiä eroja ja yhtäläisyyksiä synkronisesti eli jossakin vaiheessa samanaikaisina havaittavina ilmiöinä ja niiden suhteina. Kontrastiivinen kielentutkimus, joka sai alkunsa 1960–1970-luvuilla, on osa typologisesti vertailevaa tutkimusta, mutta se ei sinänsä kuulu suoraan kielitypologian alaan (Jørgensen 1982, 24). Yleisesti ottaen tämänkaltaista kielten vertailevaa tutkimusta ei voi pitää yhtenäisenä tutkimussuuntana vaan pikemmin tutkimusnäkökulmana, jota sovelletaan eri teorioiden ja menetelmien piirissä (ks. esim. K. Häkkinen 1987; Järventausta 2013). Kontrastiivisen tutkimuksen piirissä tähdätään myös hyvin erilaisiin päämääriin – sekä teoreettisiin että soveltaviin tarkoituksiin. Alun perin tämä suunta on lähtenyt vieraan tai toisen kielen oppimis- ja opetustarpeista, ja tutkimus on tarjonnut aineksia lähinnä virheanalyysimenetelmää varten (ks. esim. Sajavaara 1977; Johansson 2007). Nykyään sekä tutkimusmenetelmien että tutkimusaiheiden kirjo on hyvin laaja syntaktisten rakenteiden vertailusta kontrastiiviseen keskustelututkimukseen.

Typologian ja kontrastiivisen kieliopin läheiseen alueeseen kuuluu vielä komparatiivinen eli historiallis-vertaileva kielitiede, joka tähtää geneettisesti läheisten kielten synkroniseen ja diakroniseen vertailuun. Historiallis-vertailevan kielitieteen metodologiaa käyttäen verrataan vaikkapa tiettyä sanastoa etsimällä säännöllisiä äännevastaavuuksia tutkittavissa kielissä. Esimerkiksi uralilaisen kielikunnan kantakielen alkuperäisiä sanoja selvitettäessä verrataan sanoja (kognaatteja) sukulaiskielissä ja päätellään mahdolliset (sääntöperäiset) äänne- ja merkitysmuutokset. (Historiallis-vertailevasta kielitieteestä ks. Anttila 1972; K. Häkkinen 1990; J. Häkkinen 2013; Historiallinen ja vertaileva kielentutkimus tk.) Kielimuotoja ja niiden variaatioita vertaillaan myös dialektologiassa keskittyen murteisiin eli puhuttuun kieleen.

Kielten vertailu voi auttaa ymmärtämään kielten eroja ja yhtäläisyyksiä ja sitä kautta myös vertailtavia kieliä itseään. Tässä on yksinkertainen esimerkki viron ja suomen vertailusta:

1) *Karu ronib puu otsa.*
2) *Karhu kiipeää puuhun.*

Kontrastiivisen kieliopin kannalta viron ja suomen esimerkit osoittavat ainakin seuraavia eroja:

- Fonologia: *karu* (vi) – *karhu* (su). *h* on hävinnyt likvidan *r* edeltä viron kieliopin rajoitusten vuoksi.
- Syntaksi*: puu otsa* (vi) – *puuhun* (su). Virossa käytetään terminatiivista postpositiota *otsa* (’päin’, ’asti’ jne.) ja suomessa illatiivisijaa. Tyypillinen ero suomen ja viron välillä onkin, että suomen paikallissijailmauksia vastaavat virossa postpositioilmaukset. Molempien kielten neutraali sanajärjestys on se, että subjekti (*karu*, *karhu*) edeltää predikaattia ja predikaattiverbi kongruoi subjektin persoonassa ja luvussa (*ronib* ja *kiipeää* ovat molemmat yksikön 3. persoonan muotoja).
- Leksikko: *ronima* (vi) – *kiivetä* (su). Viron verbi *ronima* tarkoittaa myös ’kontata’ tai ’kaivautua läpi’ ja keskittää huomion ennen kaikkea ruumiin liikkeisiin. Suomen *kiivetä* ilmaisee puhtaammin ylöspäin suuntautuvaa vertikaalista liikettä.

**Kielioppiteorian kehittäminen**. Kieliopin teoria ei suinkaan ole vielä – jos koskaan – täysin valmis. Vanhastaan tunnettujen kategorioiden tarkempi analyysi ja kielten vertailun avulla saatu uusi tieto kategorioista ja niiden ominaisuuksista on pohjana uusille teorioille.

Esimerkkinä perinteisten kategorioiden tarkemmasta analyysista toimii vaikkapa Chomskyn (1974) ajatus, että sanaluokat substantiivi (N), verbi (V), adjektiivi (A) ja p-positio (P) voidaan esittää ”substantiivimaisuutta” ja ”verbimäisyyttä” kuvaavien piirteiden +/–N ja +/–V avulla: verbi on [–N, +V], substantiivi on [+N, –V], adjektiivi [+N, +V] ja p-positio [–N, –V]. Näin voidaan päästä yksinkertaisempaan kuvaukseen ja osoittaa sanaluokkien keskinäiset suhteet entistä paremmin, kun esimerkiksi perinteisesti nomineiksi luokitellut sanaluokat substantiivi ja adjektiivi on kuvattu samalla piirteellä +N. Chomskyn ajatusta on toki kritisoitu ja muokattu, ja piirteiden toimivuutta kaikille sanaluokille on epäilty ja arvioitu uudelleen. Esimerkki kuitenkin kuvaa hyvin sitä, että kieliopin tutkimuksessa voidaan arvioida uudelleen niinkin ikivanhojen kategorioiden kuin sanaluokkien asemaa kielijärjestelmässä.

## 3.3. Eri kielioppiteorioita ja niiden keskeisiä eroja

Kieliopin tutkimuksessa ja mallintamisessa on hyvin paljon erilaisia koulukuntia ja traditioita, emmekä pyri tässä luetteloimaan tai kuvaamaan niitä kaikkia. Esimerkkeinä erilaisista kielioppimalleista ovat vaikkapa seuraavat (suluissa tradition perustaja tai keskeinen tutkija, johon koulukunta tavallisesti yhdistetään):

Perinteinen kielioppi

Generatiivinen kielioppi (Noam Chomsky)

Kognitiivinen kielioppi (Ronald Langacker)

Muitakin suosittuja ja merkittäviä traditioita näiden lisäksi toki on. Erityisesti kielitypologit käyttävät kuvaustensa perustana ns.

**peruskielioppiteoriaa** (ks. Kielitypologian menetelmät; Kielten dokumentointi ja kieliopin kuvaus; Antropologinen lingvistiikka tk.). Seuraavan vertailun tarkoituksena on näyttää esimerkinomaisesti, minkä tyyppisiin asioihin traditioiden ja koulukuntien erot pohjautuvat, ei antaa kattavaa kuvaa kaikista kieliopin tärkeistä koulukunnista.

**Perinteinen kielioppi** on lukijoille tuttu niin sanottu koulukielioppi. Se perustuu antiikin kreikkalaiseen filosofiaan. Erityisesti lauseenjäsennys sekä sanaluokat ja osin myös muoto-opilliset kategoriat (sija, aikamuoto, tapamuoto jne.) ovat antiikin tutkijoiden kehittämiä. Perinteinen kielioppi on osoittautunut käytännölliseksi kielen opetuksessa, ja sen kategoriat ovat taustalla myös uudemmissa kieliopin malleissa. Perinteisen kieliopin ongelmiksi tieteellisen kielenkuvauksen mallina voidaan katsoa kategorioiden määritelmien yleisluontoisuus (esim. ”subjekti on lauseen tekijä”) sekä sen tapa keskittyä vain perusasioihin ja koota hankalat ilmiöt jäännösluokkiin (esim. adverbiaali, indefiniittipronomini ym.).

**Generatiivisen kieliopin** lähtölaukaukseksi katsotaan yleisesti Noam Chomskyn teos *Syntactic Structures* (1957). Uutuutena edeltävään strukturalistiseen kielioppiperinteeseen oli ajatus, että epäsäännöllisyyttä lause- ja lausekerakenteeseen aiheuttavien ilmiöiden, kuten vaihtelevien sanajärjestysten, passiivirakenteiden, kysymysrakenteiden ja ellipsien, taustalle oletetaan yksinkertaisia lausekerakennesääntöjä noudattava syvärakenne. Näin lausekerakennesäännöt pysyvät yksinkertaisina ja argumenttirakenne voidaan kuvata syvärakenteessa täydellisenä. Hintana on se, että tarvitaan mutkikas säännöstö eli transformaatiosäännöt, joilla syvärakenteesta johdetaan pintarakenteet eli ne rakenteet, jotka kielenpuhuja tunnistaa hyväksyttäväksi kieleksi (ks. Kielitaju kielentutkijan työkaluna tk.) ja jollaisia todellisissa teksteissä esiintyy. Lauseoppi on chomskylaisessa generatiivisessa kieliopissa keskeinen osa-alue: fonologinen ja semanttinen tulkinta johdetaan lauserakenteista. Chomsky (1965) on korostanut, että generatiivisen kieliopin tavoitteena on kuvata kieli sellaisena kuin se on ihmismielessä.

Chomskylaisesta generatiivisesta kielitieteestä on ajan kuluessa kehittynyt useita eri kielioppimalleja; esimerkiksi 1960- ja 1970-lukujen taitteessa generatiivinen kielioppi jakaantui generatiiviseen semantiikkaan

ja tulkitsevaan semantiikkaan. Kielioppiteorioiden erkaantuminen perustui lähinnä eri näkemyksiin siitä, mikä on kielen rakenteen ja kielen merkityksen keskinäinen suhde kielijärjestelmässä sekä kielen suhde ontologiaan: Generatiivinen semantiikka oletti merkityksen olevan ensisijainen ja pyrki liittämään lauserakenteeseen myös merkityskategorioita. Tulkitseva semantiikka lähti siitä, että rakenne (lähinnä lauserakenne) on ensisijainen merkitykseen nähden ja merkityksen kuvaus on pidettävä erillään lauserakenteen kuvauksesta. Ero on siis toisaalta kuvaustekninen, ja toisaalta se liittyy kielitieteilijän tapaan hahmottaa tutkimuskohteensa. Generatiivisesta semantiikasta kehittyi vähitellen mm. kognitiivinen kielioppi (Langacker 1987; 1991) ja tulkitsevasta semantiikasta mm. minimalismi (Chomsky 1995) ja konseptuaalinen semantiikka (Jackendoff 1983; 1990). Minimalismin ja konseptuaalisen semantiikan keskeinen ero on se, että minimalismi korostaa lauserakenteen ensisijaisuutta merkitykseen ja fonologiaan nähden, kun taas konseptuaalinen semantiikka näkee eri tasot rinnakkaisina eikä toisilleen alisteisina kielen alajärjestelminä. Suomen kielestä generatiivista tutkimusta ovat tehneet eri aikoina muiden muassa Kalevi Wiik, Fred Karlsson, Auli Hakulinen, Osmo Ikola, Urpo Nikanne, Geda Paulsen, Oksana Petrova ja Michaela Pörn.

**Kognitiivinen kielioppi** eroaa generatiivisesta kieliopista ja myös perinteisestä kieliopista monessa teoreettisesti perustavanluonteisessa suhteessa. Kognitiivisen kieliopin keskeisin kehittäjä on Ronald Langacker (1987; 1991; 2016). Teorian mukaan kielioppi on rakenteeltaan polaarinen ja perustuu toisaalta fonologiaan, toisaalta semantiikkaan. Tämä kaksinapaisuus koskee teorian mukaan kaikkia kielen perusyksiköitä, joita ovat fonologinen, semanttinen ja symbolinen yksikkö. Fonologian ja semantiikan välillä on jatkumoluonteinen suhde, ei selviä rajoja. Tähän liittyy yksi kognitiivisen kieliopin keskeinen teesi eli kieliopin selittäminen semanttisten rakenteiden konventionaalistuneen symbolisaation kautta. Kieli on Langackerin (1987, 11) mukaan luonteeltaan symbolinen muodostaen avoimen joukon kielellisiä merkkejä sekä ilmaisuja, joilla on sekä fonologinen että semanttinen representaatio. Kielen perusyksiköt sisältyvät skeemoihin eli ihmismielen tietorakenteisiin, jotka ohjaavat ja jäsentävät kokemuksia ja käsityksiä maailmasta.

Skeemat muodostavat laajempia verkostoja. Esimerkiksi kategorioiden ”hedelmä” ja ”omena” välillä vallitsee ns. täysin skemaattinen suhde (engl. *full schematicity*), kun taas ”hedelmä” ja ”tomaatti” ilmentävät osittain skemaattista suhdetta (engl. *partial schematicity*); käsitteet ”hedelmä”, ”omena” ja ”tomaatti” muodostavat yhdessä skemaattisen verkoston. (Langacker 1987, 68–75; ks. myös Leino 1993.)

Yksi tämän teorian keskeisiä käsitteitä on jatkumo: syntaksin, morfologian ja leksikon välillä ei ajatella olevan selvää rajaa, eikä kieliopin katsota koostuvan erillisistä komponenteista vaan ainoastaan jatkumoista. Leksikon katsotaan muodostavan kieliopin kanssa selkeästi erottamattoman symbolisten rakenteiden jatkumon. Lauseopin ei katsota olevan itsenäinen kieliopin osa, vaan sekin on symbolinen representaatio, joka sisältyy semanttiseen representaatioon. Semanttisella representaatiolla ei oleteta olevan syvärakennetta, eikä rakenteiden välillä ole johtosuhteita, transformaatioita. (Langacker 1987, 2–3.) Kognitiivinen kielioppi olettaa kielijärjestelmän perustuvan prototyyppisiin eli asteittaisiin kategorioihin, vastakohtana esimerkiksi generatiivisen kieliopin selvärajaisiin klassisiin kategorioihin. Toisin sanoen kognitiivisessa kieliopissa kielen kategorioiden välillä nähdään olevan asteittaisia eroja, ja kategoriat saattavat ominaisuuksiltaan kuulua sekä selviin että vähemmän selviin tapauksiin (kielitieteellisistä kategoriointiperiaatteista ks. Taylor 1989, 54–59).

Semanttinen rakenne on kognitiivisessa kieliopissa kielikohtainen; kieliopin ei oleteta olevan universaali. Semanttisia rakenteita käsitellään kognitiivisessa kieliopissa käsiterakenteen kautta. Se heijastaa konventionaalisia mielikuvia (engl. *conventional imagery*) eli sitä tapaa, miten tilanteita käsitteistetään. *Käsitteistyksellä* (engl. *conceptualization*) tarkoitetaan joko kielenkäyttäjän tulkintaa tilanteesta, jota hän kielentää, tai kieleen vakiintunutta tapaa jäsentää tilanteita. Mielikuvallisuus (usein myös termeillä *metafora* tai *figuratiivinen kieli* kutsutut ilmiöt) käsitetäänkin kieliopin keskeisenä ilmentymänä. Langackerin (1987, 110) mukaan mielikuvallisuus on ajattelu- tai ilmaisukeino, jolla konstruoimme tajutun tilanteen vaihtoehtoisina mielikuvina. Langacker (1987, 63) tekee selväksi, että kielioppi ei ole rakenteita generoiva systeemi vaan kielellisten yksiköiden varasto (”simply an inventory of linguistic units”).

Samoin kuin generatiivisen kieliopin myös kognitiivisen kieliopin tavoitteena on kuvata kieli sellaisena kuin se ihmisen mielessä on. Kognitiivinen kielioppi eroaa generatiivisesta kieliteorista siinä, että se ei sisällytä kielen rakenteeseen synnynnäisiä mekanismeja. Teoria pyrkii kuvaamaan kieltä ”psykologisesti todellisena”. Taustaoletuksena on ajatus, että kielen rakenteiden kuvaus on sellaisenaan hypoteesi kognitiivisesta representaatiosta (Langacker 1987, 56). Toisin sanoen kielioppi on olemukseltaan psykologinen representaatio, joka Langackerin (1987, 57) mukaan ei kuitenkaan ole välttämättä autonominen eikä selvästi rajattavissa eikä muusta kognitiosta erillään oleva psykologinen ilmiö.

Kielioppia tarkastelee psykologisesta ja semanttisesta näkökulmasta myös kognitiivisen kielitieteen edustaja Leonard Talmy (ks. esim. 1985; 1988; 2000), jonka näkemyksessä kieli on yksi (suhteellisen) selvästi erotettava kognitiivinen systeemi esim. havainnointi- ja päättelykyvyn, muistin, kulttuurijärjestelmän ja motorisen hallinnan ohella. Nämä systeemit eroavat esimerkiksi Fodorin (1983) selvärajaisista moduuleista siinä, että omien uniikkien ominaisuuksien lisäksi niillä voi olla rakenteellisia yhtymäkohtia joidenkin tai kaikkien muiden systeemien kanssa. (Talmy 2000, 15–16.) Talmy on kognitiivisen kieliopin tärkeä kehittäjä, ja hän on ottanut käyttöön ja kehittänyt monia nykyään laajasti sovellettavia käsitteitä, kuten *voimadynamiikan* ja *fiktiivisen liikkeen*. Hän on perinpohjainen semanttis-kognitiivisten kategorioiden erittelijä – joitakin esimerkkejä ilmiöistä, joita Talmy on analysoinut, ovat liike ja kausatiivisuus.

Suomessa kognitiivisen kieliopin ideoita on sovellettu esimerkiksi paikallissijojen ja adpositioiden kuvaamiseen, verbisemantiikkaan ja metaforatutkimukseen. Tutkijoita, jotka ovat tehneet työtä tämän tradition parissa, ovat esimerkiksi Pentti Leino, Tiina Onikki-Rantajääskö, Tuomas Huumo, Jaakko Leino, Ilona Herlin ja Jari Sivonen.

Erilaisilla kielioppiteorioilla saattaa siis olla sekä yhtymäkohtia että radikaalejakin eroja keskenään. Taulukkoon 1 on koottu edellä käsiteltyjen kielioppimallien joitain olennaisia eroja ja yhtäläisyyksiä.

Taulukko 1. Perinteisen kieliopin, generatiivisen kieliopin sekä kognitiivisen kieliopin vertailu.

| | Perinteinen kielioppi | Generatiivinen kielioppi | Kognitiivinen kielioppi |
|---|---|---|---|
| **Kognitiivinen ulottuvuus** | **Ei:** Kognitiivinen ulottuvuus ei ole keskeinen osa kuvaustraditiota. | **Kyllä:** Kieliopin laadinnan tavoitteena on kuvata ja selittää puhuja-kuulijan kielikyky. | **Kyllä:** Kielenkuvauksen tavoitteena on kuvata kieli niin, että se vastaa kielenkäyttäjän mentaalista hahmotusta |
| **Formalisointi** | **Ei:** Kuvaus perustuu kategorioihin, jotka määritellään usein tyypillisten tapausten mukaan. Esimerkiksi lauseopissa lauseen subjekti määritellään yleensä tyypillisen tapauksen mukaisesti lauseen ilmaiseman toiminnan tekijäksi. Toisaalta vaikkapa määritteiden ja pääsanojen suhteet pyritään selvittämään ja kuvaamaan formaalisti. | **Kyllä:** Formalisointi on keskeinen osa metodologiaa. | **Ei:** Formalisointi ei ole keskeinen osa kielenkuvauksen metodologiaa, mutta kuvauksia pyritään kuitenkin esittämään systemaattisella tavalla kuvioina. |
| **Kielijärjestelmän osa-alueiden erillisyys (modulaarisuus)** | **Kyllä:** Kieliopin eri osa-alueet (äänneoppi, lauseoppi, muoto-oppi). | **Kyllä:** Modulaarisuus on keskeinen osa teoriaa. | **Ei:** Modulaarisuuden kieltäminen on osa teoriaa. |
| **Kielen ”tärkein” osa-alue tai teoriaperinteen fokusointikohde** | **Ei selvää fokusaluetta:** Äänneoppia, muoto-oppia ja lauseoppia on kuvattu erikseen. | **Syntaksi** on katsottu keskeiseksi osaksi kielen rakennetta. | **Semantiikka** on keskeinen osa kieltä, koska merkityksen katsotaan olevan myös muodon perusta. |
| **Universaalisuus vs. relatiivisuus** | **Ei selvästi kumpikaan.** Kieliä on kuvattu klassisten kielten mallin mukaisesti, mikä on eräänlaista universalismia. Tämä ei kuitenkaan ole teorianmuodostuksen keskeinen osa vaan perustuu perinteeseen ja käytännön tarpeisiin. Lisäksi klassisten kielten kielioppeja on käytetty mallina kaikkien kielten kuvauksessa. | Kielellinen **universalismi:** Luonnollisilla kielillä oletetaan olevan yhteinen kielijärjestelmän toteutumista säätelevä osa. | Kielellinen **relativismi**: Kielten välisten samankaltaisuuksien katsotaan perustuvan muihin kuin kielellisiin tekijöihin, esimerkiksi siihen, että ihmiskeho toimii kaikilla kielenpuhujilla samalla lailla. |

Vaikka edellä on käsitelty kolmea traditiota, on hyvä muistuttaa lukijaa siitä, että kielioppitraditioita on paljon. Mainittakoon tässä vielä vaikkapa M. A. K. Hallidayn (1994) ajatuksiin perustuva **systeemis-funktionaalinen kielioppi**, jossa kielen rakenne ja lauseen merkitys jaetaan kolmeen tasoon: rakenteen tasoon (lausekerakenne), syntaktisten funktioiden tasoon (lauseenjäsenet) ja semanttisten funktioiden tasoon (argumenttirakenne). Nämä kolme tasoa toimivat samanaikaisesti, ja ne kuvataan kukin erikseen. Systeemis-funktionaalinen kielioppi painottaa mentaalisten prosessien sijasta kontekstin merkitystä kielelliselle rakenteelle. Teoria tutkii vuorovaikutuksen roolia kielen käytössä kysyen, miten kieltä käytetään tiettyjen tavoitteiden saavuttamiseksi. Nimityksen toinen osa, *funktionaalinen*, viittaa kielenkäytön ensisijaisuuteen. Nimityksen ensimmäinen puoli, *systeeminen*, puolestaan tulee siitä, että SF-kieliopissa tarkastellaan myös syntagmaattisia suhteita eli kielenyksiköiden suhteita toisiin yksiköihin, joiden kanssa ne esiintyvät peräkkäin samassa jaksossa. Suomessa alan tutkijoita ovat esimerkiksi Susanna Shore, Katarzyna Wojciechowicz, Vesa Heikkinen, Eija Ventola ja Yrjö Lauranto.

## 3.4. Kielioppiteorian soveltaminen

Kielioppia on käytetty vuosituhanten aikana erityisesti kielten opettamiseen sekä kielenhuoltoon. Vieraita kieliä on kuvattu kielioppien avulla, ja kieliopin sääntöjä on käytetty opetuksessa yhdessä sanakirjojen kanssa kielijärjestelmän opettamiseen. Kielenhuoltoon kielioppia on käytetty siten, että kieliopin sääntö on tulkittu ohjeeksi oikeasta rakenteesta eikä ainoastaan kielen kuvaukseksi. Esimerkiksi suomen kielen possessiivisuffiksit jäävät puhekielessä useimmiten pois, mutta kirjoitetussa kielessä niitä pitää käyttää erityisesti silloin, kun persoonapronomini toimii genetiiviattribuuttina: ***minun** kirjani*, ***sinun** kirjasi*, ***hänen** kirjansa*. Kirjoitetussa asiatekstissä possessiivisuffiksin pois jättäminen olisi virhe. Kielenhuollolla hillitään kielellisten rakenteiden variaatiota asiatyylisissä teksteissä.

Vaikka leksikografiassa päähuomio on sanoilla, tieto kuvattavan kielen kieliopista on leksikografeille välttämätön. Kuten jo johdantoluvussa

todettiin, sanaston ja kieliopin välinen raja ei ole selvä, ja niinpä myöskään kielioppi ja sanakirja eivät ole jyrkästi erotettavissa (Kytömäki 1991). Jokainen sana sisältää tietoa sen funktiosta laajemmassa ympäristössä, ja sanat saattavat olla esimerkiksi johtamisprosessin tulosta. Myös sanojen kategorisointi sekä näkökulma siitä, mikä kuuluu sanakirjaan ja mikä ei, riippuu teoreettisesta lähestymistavasta. Sanakirjan kirjoittajalle tämä on olennainen käytännön kysymys. Jos lähdetään vaikka konstruktiokieliopin[2] esittämästä oletuksesta, että sekä sanat että sitä laajemmat yhdistelmät (esimerkiksi transitiivilause, kollokaatiot, idiomit jne.) ovat tallentuneet kielenkäyttäjän mieleen periaatteessa samantyyppisinä muodon ja funktion pareina, niin sekä perinteiset sanat että niiden konstellaatiot tulisi listata sanakirjaan. Tällä tavalla sanakirjat voisivat jopa ”nielaista” kieliopit, kuten konstruktiokieliopin radikaalein edustaja William Croft yhdessä Logan Suttonin (2014) kanssa toteaa.

Kieliopin tutkimusta sovelletaan opetuksen, kielenhuollon ja sanakirjatyön lisäksi moniin muihinkin tarkoituksiin. Esimerkiksi kieliteknologiset sovellukset ovat nykyisin nopeasti kehittyvä ala, johon kuuluu hyvin erilaisia sekä puhetta että tekstejä käsittäviä alueita aina äänipohjaisista dialogijärjestelmistä konekäännös- tai kielenopetusohjelmiin. Kieliopin tutkimusta tarvitaan lisäksi muun muassa uhanalaisten kielten elvytyshankkeissa (ks. Antropologinen lingvistiikka tk.). Tietoa kieliopista tarvitaan myös kielellis-kognitiivisten oireiden tutkimuksessa ja hoidossa, esimerkiksi neuropsykologiassa ja logopediassa (ks. Psyko- ja neurolingvistiikka tk.).

## 3.5. Tutkimuksen kulku

Kieliopillinen tutkimus etenee havaintojen tekemisestä teorianmuodostukseen tutkimuksen tavoitteiden mukaisesti. Tutkimusperinteen vakiintuneisuuden ansiosta yleensä kielen havainnoinnissa voidaan käyttää jo vakiintuneita kieliopin kategorioita. Havainnointia voidaan tehdä eri tavoin. Esimerkiksi:

1. Voidaan tunnistaa jostain kieliaineistosta (esimerkiksi nauhoituksesta tai kirjoituksesta) tutkimuksen kohteena olevia ilmiöitä (vaikkapa sijamuotoja, pronomineja, tavuja, nasaaliäänteitä tai muita kieliopin tutkimuksen kategorioita) ja selvittää tarkemmin niiden esiintymistä. Näin saadaan ilmiöistä tarkempaa tietoa ja voidaan selvittää, kuinka tarkkailtava kategoria toimii tutkitussa kielessä, sekä myös tarkentaa aiempia oletuksia: Onko kyseisessä kielessä tätä kategoriaa (onko esim. suomen kielen äänneoppiin oletettava piirre [soinnillinen], vaikka sitä ei tarvita suomen äännejärjestelmän yksiköiden erottamiseen eikä se siis ole suomen kielessä distinktiivinen)? Pitääkö kielessä jotkin kategoriat yhdistää (pitäisikö esim. suomen muoto-opissa kategoriat tempus ja modus yhdistää, kun niitä ilmaisevat tunnukset eivät esiinny koskaan samassa sananmuodossa) vai pitääkö olettaa joitain kokonaan uusia kategorioita (esimerkiksi jakamaan suomen kielen indefiniittipronominien sekavaa joukkoa)?
2. Havainnoinnin kohteena voivat olla eri kielten kielioppien valmiit kuvaukset. Tämä on tyypillistä erityisesti typologisessa kielentutkimuksessa, jossa selvitetään kielen rakenteiden ilmenemistä maailman kielissä (ks. Kielitypologian menetelmät tk.).
3. Kun tunnetaan kielen rakenne aiemman tutkimuksen pohjalta, voidaan laatia testilauseita, joiden ominaisuudet tunnetaan, ja koettaa, miten jonkin muuttujan vaihtaminen vaikuttaa kielenpuhujan tulkintaan (aiheesta enemmän tämän teoksen luvussa Kielitaju kielentutkijan työkaluna).

Kun aineisto on hankittu, pyritään sovittamaan havainnot kieliopin kuvaukseen niin, että uusi kieliopin kuvaus on järjestelmä, joka vastaa aineistosta saatuja havaintoja ja osoittaa, mikä on mahdollinen rakenne ja mikä ei. Kieliopillisten kuvausten keskinäisen paremmuuden arviointi on yksi tutkimuskohde, ja sen tärkeyttä on generatiivisessa kieliopissa korostanut erityisesti Chomsky (1965).

## 3.6. Kieliopin suhteet muihin kielen osa-alueisiin ja niiden tutkimukseen

Kieliopin tutkimus käsittelee kielen keskeistä osaa, rakenteen muodostamaa järjestelmää, ja on siinä mielessä kielentutkimuksen keskiössä. Sen vuoksi yhteydet muihin kielen osa-alueisiin ovat kiinteät eikä aina ole selvää, mitkä ilmiöt pitää laskea osaksi kielioppia ja mitkä jätetään kieliopin ulkopuolelle. Selvää luonnollista rajaa ei olekaan, erityisesti, kun pyritään erottamaan vaikkapa seuraavat kielen osa-alueet:

**Sanasto**: Sanat muodostuvat äänteistä, ne esiintyvät aina jossain muodossa ja liittyvät toisiinsa lauseiksi ja lausekkeiksi. Niinpä sanasto voidaan myös laskea osaksi kielioppia, koska äänneoppia, muoto-oppia ja lauseoppia ei voi soveltaa ilman sanoja. Perinteisesti on tehty jako kielioppiin ja sanastoon, mutta kielijärjestelmää ei voi kuvata ilman sanojen kuvausta.

**Merkitys**: Kieliopin tutkimusta ei voi jyrkästi erottaa semantiikasta, koska kieliopin jokainen osa on kytköksissä merkitykseen. Foneemit on perinteisesti määritelty pienimmiksi merkitystä erottaviksi yksiköiksi, ja kielen foneemit perustellaan minimipareilla (esimerkiksi suomen likvidat [l] ja [r] ovat eri foneemeja, koska on sellaisia minimipareja kuin esim. *kaula* – *kaura* ja ***l**epäisin* – ***r**epäisin*). Muoto-opin muodoilla on yleensä jokin oma tyypillinen merkitysalueensa: tempukset kuvaavat aikaa, monikon tunnus useutta jne. Lauserakenteen muutos voi myös vaikuttaa lauseen merkitykseen, esimerkiksi korosteisuuteen (vrt. *Koululaisille tarjotaan näkkileipää* [neutraali korostus] – *Näkkileipää koululaisille tarjotaan* [*näkkileipä* on korostetussa asemassa: nimenomaan näkkileipää eikä jotain muuta!]).

**Kielen käyttö ja vuorovaikutustutkimus**: Koska kieli on ensisijaisesti ihmisten viestinnän ja vuorovaikutuksen väline, kieliopin on sopeudutettava kielen käyttötilanteisiin. Siksi on tärkeää selvittää, kuinka kielen rakenne toimii eri tilanteissa. (Aiheesta enemmän tämän teoksen luvussa Keskustelunanalyysi ja vuorovaikutuslingvistiikka.)

# 4. Yhteenveto

Tässä artikkelissa on käsitelty kieliopin tutkimusta. Tiivistetysti artikkelin keskeiset huomiot ovat seuraavat:

1. Kieliopin tutkimus on kielijärjestelmän tutkimusta rakenteen osalta. Kieliopin tutkimuksen ja kielioppien laatiminen edellyttää sitä, että kielessä on ainakin jonkin verran sellaista säännönmukaisuutta, että siitä voidaan laatia kielioppi. Kieliopin tutkimus perustuu säännölliseen havainnointiin ja loogiseen päättelyyn kuten muukin tieteellinen tutkimus. Tutkimuksen tavoitteena on luoda mahdollisimman kattavia ja elegantteja kuvauksia yksittäisten kielten rakenteiden järjestelmistä sekä kielten välisistä yhtäläisyyksistä ja eroista.
2. Eri traditiot poikkeavat toisistaan niin kuvauksen kohteen määrittelyssä kuin kielijärjestelmän luonteen hahmotuksessa tai menetelmässä, jolla järjestelmää tutkitaan ja kuvataan.
3. Kielioppeja on tarvittu muun muassa kielten opetuksen sekä oikeakielisyyden tarpeisiin. Niinpä kelioppien laatimisella onkin tuhansien vuosien perinne, ja myös suomen kielen kielioppeja on kirjoitettu jo neljänsadan vuoden ajan.
4. Kieliopin alaan kuuluvien ilmiöiden rajaaminen ei ole itsestään selvää tai perustu sellaisenaan havaintoon: rajan vetäminen kieliopin ja semantiikan tai kieliopin ja leksikon väliin edellyttää teoreettista näkemystä koko kielijärjestelmästä ja siksi eri tutkimustraditiot ja kielioppimallit vetävät rajan eri kohtiin – tai ovat vetämättä rajaa lainkaan.

**Aiheesta lisää:**


Itkonen, Esa. 1991. *Universal History of Linguistics: India, China, Arabia, Europe*. Amsterdam: John Benjamins.

Karlsson, Fred. 2015 [1994]. *Yleinen kielitiede*. (5. p.). Helsinki: Gaudeamus.

Strazny, Philipp (toim.). 2005. *Encyclopedia of Linguistics*. London: Routledge.

## VIITTEET

1 Rättyä (2014) sisällyttää kielitiedon alaan myös oikeakielisyyteen ohjaavan eli preskriptiivisen kieliopin "metakielen", esimerkiksi kirjakielen oikeinkirjoitusohjeistukset.

2 Konstruktiokielioppi on suuntaus, jonka mukaan kieliopin pohjana ovat konventionaaliset muodon ja merkityksen yhteenliittymät. Eri kuvausmalleissa *konstruktion* käsite ja sen asema yleisessä kielijärjestelmässä vaihtelevat jonkin verran (ks. esim. Fillmore 1988; Goldberg 1995; Croft 2001; Östman & Fried 2004).

## KIRJALLISUUS


Allan, Keith (toim.). 2013. *Oxford Handbook of History of Linguistics*. Oxford: Oxford University Press.

Anttila, Raimo. 1972. *An Introduction to Historical and Comparative Linguistics*. New York (NY): Macmillan.

Aristoteles. 1994. *Kategoriat*. [*Katēgoriai*.] Teokset 1. Suomentanut ja selitykset laatinut Lauri Carlson. Helsinki: Gaudeamus.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge (MA): MIT Press.

Chomsky, Noam. 1974. "The Amherst Lectures". Lectures given at the 1974 Linguistic Institute, University of Massachusetts. Amherst: Université de Paris VII.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. Westport, Connecticut: Praeger Publishers.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge (MA): The MIT Press.

Chomsky, Noam. 2000. *New Horizons in the Study of Language and Mind*. Cambridge (MA): Cambridge University Press.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press: Oxford.

Croft, William & Sutton, Logan. 2014. Construction grammar and lexicography. Julkaisussa: Hanks, Patrick & de Schryver, Gilles-Maurice (toim.) *International Handbook of Lexis and Lexicography*. New York (NY): Springer.

Dryer, Matthew S. & Haspelmath, Martin (toim.). 2013. The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. [verkkoaineisto]. [viitattu 22.4.2018]. Saatavissa: http://wals.info.

Fillmore, Charles J. 1988. The mechanisms of 'Construction Grammar'. *Berkeley Linguistic Society* 14, 35–55.

Fodor, Jerry. 1983. *The Modularity of Mind: an Essay on Faculty Psychology*. Cambridge (MA): MIT Press.

Foley, William A. & Valin Jr., Robert D. van. 1984. *Functional Syntax and Universal Grammar*. Cambridge Studies in Linguistics, Vol. 38. Cambridge: University Press.

Goldberg, Adele. 1995. *Constructions*. Chicago (IL): University of Chicago Press.

Hakulinen, Auli, Kauppinen, Anneli, Leiwo, Matti, Paunonen, Heikki, Räikkälä, Anneli, Saukkonen, Pauli, Yli-Vakkuri, Valma, Östman, Jan-Ola & Alho, Irja (toim.). 1996. *Kieli ja sen kieliopit. Opetuksen suuntaviivoja*. Helsinki: Opetusministeriö ja Painatuskeskus.

Hakulinen, Auli, Vilkuna, Maria, Korhonen, Riitta, Koivisto, Vesa, Heinonen, Tarja Riitta & Alho, Irja. 2004. *Iso suomen kielioppi*. SKS:n toimituksia (SKST) 950. Helsinki: Suomalaisen Kirjallisuuden Seura.

Halliday, M. A. K. 1994. *Introduction to Functional Grammar.* (2. p.). London: Edward Arnold.

Harjunpää, Mira. 2016. Nominilausekkeiden kuvaus latinan kieliopeissa myöhäisantiikista humanistirenessanssiin. Pro gradu -tutkielma, Helsingin yliopisto.

Householder, Fred W. 1981. *The Syntax of Apollonius Dyscolus*. Studies in the History of Linguistics, 23. Amsterdam: John Benjamins.

Häkkinen, Jaakko. 2013. Kielen muutos. Julkaisussa: Kolehmainen, Leena, Miestamo, Matti & Nordlund, Taru (toim.) *Kielten vertailun metodiikka*. Helsinki: Suomalaisen Kirjallisuuden Seura, 171–218.

Häkkinen, Kaisa. 1987. Kontrastiivisesta tutkimuksesta. Julkaisussa: Koski, Mauno (toim.) *Kontrastiivista kielentutkimusta* 1. Turku: Åbo Akademi, 5–25.

Häkkinen, Kaisa. 1990. *Mistä sanat tulevat. Suomalaista etymologiaa*. Tietolipas 117. Helsinki: Suomalaisen Kirjallisuuden Seura.

Ikola, Osmo. 1964. Verbisyntaksia. *Lauseopin kysymyksiä. Tutkielmia nykysuomen syntaksin alalta*. (2. p.). Helsinki: Suomalaisen Kirjallisuuden Seura.

Ikola, Osmo. 1974. *Lauseenvastikeoppia*. Tietolipas 76. Helsinki: Suomalaisen Kirjallisuuden Seura.

Itkonen, Esa. 1978. *Grammatical Theory and Metascience: A Critical Investigation into the Philosophical and Methodological Foundations of 'Autonomous' Linguistics*. Amsterdam: John Benjamins.

Itkonen, Esa. 1991. *Universal History of Linguistics: India, China, Arabia, Europe*. Amsterdam: John Benjamins.

Jackendoff, Ray. 1983. *Semantics and Cognition*. Cambridge (MA): MIT Press.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge (MA): MIT Press.

Johansson, Stig. 2007. *Seeing through Multilingual Corpora. On the Use of Corpora in Contrastive Studies*. Amsterdam: John Benjamins.

Järventausta, Marja. 2013. Kontrastiivinen tutkimus vertailevan kielentutkimuksen kentässä. Julkaisussa: Kolehmainen, Leena, Miestamo, Matti & Nordlund, Taru (toim.) *Kielten vertailun metodiikka*. Helsinki: Suomalaisen Kirjallisuuden Seura, 96–134.

Jørgensen, Normann J. 1982. *Contrastive Linguistics. What? How and Why? Where and When?* Anglica et Americana 15. Copenhagen: University of Copenhagen.

Kalliokoski, Jyrki, Kumenius, Johanna, Luukka, Minna-Riitta, Mustaparta, Anna-Kaisa, Nissilä, Leena & Tuomi, Marja. 2015. Kielitiedon opettamisesta. Julkaisussa: Mustaparta, Anna-Kaisa (toim.) *Kieli koulun ytimessä – näkökulmia kielikasvatukseen*. Helsinki: Opetushallitus. [verkkoaineisto]. [viitattu 22.4.2018]. Saatavissa: https://www.oph.fi/fi/tilastot-ja-julkaisut/julkaisut/kieli-koulun-ytimessa-nakokulmia-kielikasvatukseen.

Karlsson, Fred. 1997. *Yleinen kielitiede Suomessa kautta aikojen*. Helsingin yliopiston yleisen kielitieteen laitoksen julkaisuja 28. Helsinki: Helsingin yliopisto.

Kolehmainen, Leena, Miestamo, Matti & Nordlund, Taru (toim.). 2013. *Kielten vertailun metodiikka*. Suomalaisen Kirjallisuuden Seuran Toimituksia 1387. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kytömäki, Leena. 1991. Mikä on johdosten paikka? *Sananjalka* 33, 7–29.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar, vol. 1: Theoretical prerequisites*. Stanford (CA): Stanford University Press.

Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar, vol. 2: Descriptive application*. Stanford (CA): Stanford University Press.

Langacker, Ronald W. 2016. *Nominal Structure in Cognitive Grammar. The Lublin Lectures*. Toimittanut Głaz, Adam, Kowalewski, Hubert & Łozowski, Przemysław. Lublin: Maria Curie-Skłodowska University Press.

Law, Vivien. 1997. Late Latin grammars in the early Middle Ages: A typological history. Julkaisussa: Law, Vivien (toim.) *Grammar and Grammarians in the Early Middle Ages*. London: Longman, 54–69.

Lehmann, Christian. 2001. Language documentation: a program. Julkaisussa: Bisang, Walter (toim.) *Aspects of Typology and Universals*. Berlin: Akademie Verlag, 83–97.

Leino, Pentti. 1993. *Suomen kielen kognitiivista kielioppia* 1. *Polysemia – kielen moniselitteisyys*. Kieli 7. Helsinki: Yliopistopaino.

Luhtala, Anneli. 2013. Pedagogical grammars before the eighteenth century. Julkaisussa: Allan, Keith (toim.) *Oxford Handbook of History of Linguistics*. Oxford: Oxford University Press, 360–378.

Rasinkangas, Reijo. 2018. *Tieteen historia*. [verkkoaineisto]. [viitattu 22.4.2018]. Saatavissa: http://www.rasinkangas.fi.

Rättyä, Kaisu. 2014. Opettajan pedagoginen sisältötieto kielitiedon opetuksen näkökulmasta. *Kieli, koulutus ja yhteiskunta* 5:3. [verkkoaineisto]. [viitattu 22.4.2018]. Saatavissa: https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-toukokuu-2014/opettajan-pedagoginen-sisaltotieto-kielitiedon-opetuksen-nakokulmasta.

Sajavaara, Kari. 1977. Contrastive linguistics past and present and a communicative approach. Julkaisussa: Sajavaara, Kari & Lehtonen, Jaakko (toim.) *Contrastive Papers 4*. Jyväskylä: Department of English, University of Jyväskylä, 9–30.

Siro, Paavo. 1964. *Suomen kielen lauseoppi*. Helsinki: Tietosanakirja Oy.

Strazny, Philipp (toim.). 2005. *Encyclopedia of Linguistics*. London: Routledge.

Talmy, Leonard. 1985. Force dynamics in language and thought. Julkaisussa: Eilfort, William H., Kroeber, Paul D. & Peterson, Karen L. (toim.) *Papers from the Parasession on Causatives and Agentivity at the 21st Regional Meeting of the Chicago Linguistic Society 21*. Chicago (IL): Chicago Linguistic Society, 293–337.

Talmy, Leonard. 1988. Force dynamics in language and cognition. *Cognitive Science* 2, 49–100.

Talmy, Leonard. 2000. *Toward a Cognitive Semantics I-II*. Cambridge (MA): MIT Press.

Taylor, John R. 1989. *Linguistic Categorization: Prototypes in Linguistic Theory*. Oxford: Clarendon Press.

Viljamaa, Toivo. 2007. *Antiikin kielitiede. Luentoihin ja esitelmiin perustuva kirjoitus*. [verkkoaineisto]. [viitattu 22.4.2018]. Saatavissa: http://docplayer.fi/8475475-Antiikin-kielitiede-teoreettisia-kiinnekohtia-toivo-viljamaa.html.

VISK = Hakulinen, Auli, Vilkuna, Maria, Korhonen, Riitta, Koivisto, Vesa, Heinonen, Tarja Riitta & Alho, Irja. 2004. *Iso suomen kielioppi*. Kotimaisten kielten tutkimuskeskuksen verkkojulkaisuja 5. [verkkoaineisto]. [viitattu 28.2.2019]. Saatavissa: http://scripta.kotus.fi/visk.

Östman, Jan-Ola & Fried, Mirjam. 2004. Historical and intellectual background of Construction Grammar. Julkaisussa: Fried, Mirjam (toim.) *Construction Grammar in a Cross-Language Perspective*. Constructional approaches to language, vol. 2. Amsterdam: John Benjamins, 1–10.

# Luku 18 Näkökulmia filologiseen tutkimukseen

*Minna Seppänen*
https://orcid.org/0000-0001-6623-5708
*Janne Skaffari*
https://orcid.org/0000-0001-7349-9904

> ”To put it crudely: it's got philology in there, whatever the label says.” (Nikolaus Wegmann)

**Mikä?**

*Filologisiksi menetelmiksi* kutsutaan perinteisiä tekstintutkimuksen menetelmiä, joissa ydinajatuksena on vahvaan kielitaitoon ja tekstin syntykontekstin tuntemukseen nojaten ymmärtää, mitä tekstissä oikeastaan sanotaan. Filologian keskiössä ovat vanhat tekstit, eikä se aina ole vain kieleen kohdistuvaa tutkimusta kapeasti ymmärrettynä.

**Katso myös:**

Luku 11 Historiallinen korpuslingvistiikka

Luku 15 Historiallinen ja vertaileva kielentutkimus

Luku 20 Tekstintutkimus

# 1. Johdanto

Tässä artikkelissa pohdimme, millaisia merkityksiä *filologialle* on annettu ja edelleen annetaan ja millaista on filologisen tutkimuksen tekeminen. Käsittelemme lyhyesti myös filologian suhdetta sen lähitutkimusaloihin. Lähdemme siitä (kenties optimistisesta) olettamuksesta, että filologian olemuksen voi ylipäätään tiivistää tämän kirjan tarpeisiin, vaikka eklektisyys menetelmien ja kysymyksenasetteluiden suhteen on alalla yleistä, mikä johtaisi pikemminkin pluralisoivaan tulkintaan. Käsittelemme tässä artikkelissa historiallisia seikkoja laajemmin kuin kenties missään muussa tämän teoksen artikkelissa; tämä on helposti puolustettavissa sillä, että filologian historia on paitsi hämmästyttävän pitkä, noin 2 300 vuotta, myös ylipäätään kaikenlaisen kielentutkimuksen historiaa.

Yleisesti voidaan sanoa, että filologisella tutkimuksella, jonka kohteena ovat laveasti määrittäen vanhat tekstit, on merkittävä rooli sillanrakentajana menneen ja nykyisen välillä. Tekstien ymmärrettäväksi tekeminen uusintaa sukupolvi toisensa jälkeen mahdollisuutta tarkastella ihmisyyttä ja siihen liittyviä asioita menneisyydessä. Verrattuna mihin tahansa muuhun säilyneeseen materiaaliin tekstien näistä asioista tarjoama kuva on rikkain ja eksplisiittisin.

Aluksi on syytä kysyä, mitä *filologialla* oikeastaan tarkoitetaan. Vastaus tähän riippuu siitä aikakaudesta, jota koskevaksi kysymys määritellään. Sanan etymologiasta on hyvä aloittaa. Kreikan *filologos* on kirjaimellisesti 'logosta rakastava'. *Logos* puolestaan voi viitata kieleen ja eritasoisiin ilmiöihin kielessä, kuten lauseeseen, argumenttiin tai sanaan. Kaikkein vanhimmissa lähteissä filologi onkin ylipäätään puhumista tai puheita rakastava ihminen ja filologia siis rakkautta, joka kohdistuu argumentointiin tai kieleen ja kirjallisuuteen yleensä. Antiikin maailmassa työtä, jonka tunnistamme filologin työksi – siis editioiden (ks. alla) ja kommentaarien laatimista sekä tekstin ilmiöiden tutkimista – tekivät ns. grammaatikot (kr. *grammatikoi*, lat. *grammatici*). Sanalla *filologia* tuohon työhön on viitattu vasta 1500-luvun loppupuolelta lähtien. Ennen 1800-lukua filologialla viitattiin kaikenlaiseen kieleen liittyvään

tutkimukseen, joskin ennen kaikkea sellaiseen, johon liittyi jokin tekstuaalinen ongelma, kuten tyypillisesti muinaiskielillä kirjoitettuun materiaaliin, oli se sitten kreikkaa, hepreaa tai sanskritia. Ennen *kielitieteen* tai *lingvistiikan* käsitteen syntyä 1800-luvulla filologiaan kuuluivat niin yksittäisten kielten ja niiden historian tutkimus kuin kielen alkuperään ja luonteeseen liittyvät pohdinnat. Läheisesti filologiseen tutkimukseen liittyivät myös retoriikan, historian ja kulttuurin tutkimus sekä menneisyyden materiaalisten jäänteiden tutkimus – esimerkiksi historia ja arkeologia ovat edelleen läheisiä ja tärkeitä monille filologeille. Kielitieteen syntymisen myötä voidaan filologian katsoa menettäneen asemiaan. Vastakkainasettelu tiivistyy 1800-luvun puolivälissä vaikuttaneen August Schleicherin muotoiluun lingvistiikasta ihmistä koskevana ”luonnontieteenä” (sa. *Naturwissenschaft*) ja filologiasta jonkinlaisena ”tulkintatieteenä” tai ”hengentieteenä” (sa. *Geisteswissenschaft*). Käsitteiden tasolla eroa ei välttämättä kuitenkaan tehty: *filologia* jäi terminä elämään monessa yhteydessä viitaten kielentutkimukseen yleensä.

# 2. Filologia ennen ja nyt

## 2.1. Määritelmiä ja lähtökohtia

Filologia on vuosisatojen aikana määritelty monella eri tavalla. Tavallisimmin filologiasta puhutaan nykyisin klassillisten kielten yhteydessä, jossa kaikki tutkittava materiaali on luonnollisesti vanhaa, vaikkakaan ei välttämättä peräisin antiikista; myös esimerkiksi Bysantin kreikkaa ja keskiajan jälkeen kirjoitettua, suurelta osin tieteellistä uuslatinaa tutkitaan. Myös muiden kielten yhteydessä filologia yhdistetään vanhoihin aineistoihin ja kielihistoriaan, mutta diakronisen, kielen muuttumista kuvaavan tutkimuksen sijaan filologinen tutkimus on usein synkronista, eli se keskittyy tietyn hetken tai kauden tilaan – vaikkapa yhden tekstin tai tekstistön ilmiöiden kuvaamiseen ja selittämiseen – eikä kielimuodon kehittymiseen ajan myötä. Kielestä riippumatta keskeistä on se, että filologia tieteellisenä lähestymistapana keskittyy teksteihin. Se ei rajaa

kiinnostustaan ainoastaan kirjallisuuteen perinteisessä mielessä: filologit tutkivat kaikenlaista kirjoitettua materiaalia genreen, säilymisalustaan tai kirjoittajaan katsomatta.

Filologia on myös toisinaan määritelty kulttuurin tutkimiseksi kielen avulla (ks. esim. Anttila 1989; Altschul 2010). Vanhastaan filologisessa mielessä kiinnostavia ovatkin olleet paitsi eritasoiset kieleen ja kirjallisuuteen liittyvät kysymykset (lauseoppi, leksikologia, retoriset keinot, tekstilajit jne.) myös muunlaiset kulttuuriin liittyvät kysymykset, kuten yhteiskunnalliset instituutiot, tavat, uskonto, oppihistoria jne. Kulttuurin tutkimiseen viittaava määritelmä siis tuntuu periaatteessa oikealta, vaikkakin lopullinen tarkkuus riippuu *kulttuurin* käsitteen ekstensiosta. Joka tapauksessa tämän kaltainen määritelmä saattaa kertoa jotain filologisesta kielikäsityksestä: kieli eri rekistereissään (joista prototyyppinen filologia on ollut kiinnostunut etenkin runokielestä) kantaa mukanaan sitä käyttänyttä yhteisöä ja kulttuuria. Kielen ja kulttuurin suhde on kompleksinen: kieli on samanaikaisesti kulttuurin ilmaus ja osa sitä.

Historiansa suhteen filologia ei oikeastaan lainkaan kuulu samaan sarjaan muiden tässä kirjassa esiteltyjen kielentutkimuksen osa-alueiden kanssa. Tämä koskee sekä itse sanan historiaa että filologiaksi käsittämämme tieteenalan historiaa: käsitteen tasolla ymmärrettynä filologian historia on moninkertaisesti pidempi kuin institutionaalisesti ymmärrettynä. Siinä missä vaikkapa *nimistöntutkimus* ja *diskurssianalyysi* sanoina pyrkivät kuvaamaan sitä, mitä näitä harjoittava tutkija tosiasiassa tekee, *filologia* tyytyy nimessään ylimalkaisen maalailevaan ”kielen rakastamiseen”. Tämä ei varsinaisesti kuulosta tieteelliseltä lähtökohdalta, vaikka voikin monen filologin tai yleisemminkin kielentutkijan kohdalla olla totta: kertoohan se asenteesta, jota voidaan luonnehtia kieltä arvostavaksi kiinnostukseksi ja pyrkimykseksi sen syvälliseen ymmärtämiseen. Ehkäpä tieteenalan nimityksen epätäsmällisyyden vuoksi myös monet luonnehdinnat filologiasta ovat merkillisen arvolatautuneita todellisen substanssin määrittelyn kustannuksella – näin esimerkiksi toteaa Hans Ulrich Gumbrecht (2003, 4): ”philology has cultivated its self-image as a patient craft whose key values are sobriety, objectivity, and rationality”.

Kuten todettua, tunnemme filologian sanana jo antiikista, samoin tieteenalan, vaikkakaan termi ja tarkoite eivät välttämättä langenneetkaan

yhteen vielä tuolloin. Länsimaisessa kulttuuripiirissä filologian voidaan katsoa saaneen alkunsa Aleksandriassa 200-luvun alkupuolella ennen ajanlaskun alkua, kreikkalaisten Ptolemaiosten hallitsemassa Egyptissä.[1] Aleksandriaan kerättiin määrätietoisella työllä kreikkalaisen kirjallisuuden kokoelmaa ja luotiin lisäksi tutkimuksen mahdollistava instituutio, jota kutsuttiin nimellä Museion. Ensisijaisen kiinnostuksen kohteena olivat homeeriset tekstit, joiden mahdollisimman alkuperäiseen ja luotettavaan muotoon saattamista varten kehittyivät tekstikritiikin menetelmät. Kovin tavaton ei ole väite, jonka mukaan homeerinen runous itsessään sisältää eräänlaista filologista tarkastelua, ja että tätä runoutta esittäneet ammattimaiset rapsodit eli runonlaulajat arkaaisen kauden kreikkalaisessa maailmassa olivat itse asiassa ensimmäisiä filologian harjoittajia, koska he olivat perehtyneitä esittämiensä laulujen sisältöön ja kieleen ja pystyivät tarvittaessa myös selittämään niitä. Homeerisessa runoudessa oli kyse kreikan kielen erityisestä variantista, joka runomitan suhteellisen tiukkoihin vaatimuksiin mukautuneena ei edustanut sinänsä mitään koskaan käytössä ollutta kielimuotoa vaan oli omalakista taidekieltä, jossa sekoittui erilaisia elementtejä sekä eri aikakausilta että eri puolilta itäistä Välimerta.

## 2.2. Status quo

Suomalaisissa yliopistoissa on vielä vuonna 2020 yllättävän monta oppiainetta, joiden nimissä sana *filologia* on edelleen jäljellä. Tämä heijastelee sanan vanhaa ja pitkäaikaista merkitystä ’kieleen kohdistuva tutkimus’: esim. englantilainen, espanjalainen, germaaninen, italialainen, kreikkalainen, latinalainen, pohjoismainen, ranskalainen ja portugalilainen filologia ovat oppiaineiden nimiä – tosin tilanne elää jatkuvasti. Joskus törmää termiin *uusfilologia* (esim. Uusfilologinen yhdistys, perustettu v. 1887), joka on luotu aikoinaan klassillisesta filologiasta erottautumisen vuoksi käsittämään ennen kaikkea muiden eurooppalaisten kielten tutkimusta. *Fennistisestä filologiasta* ei tavallisesti puhuta (ks. kuitenkin Uusitalo 2019); myös suomalaisen kirjallisuuden tutkimus yliopistolaitoksissamme on paljolti eriytynyt kielen- ja

tekstintutkimuksesta. Muiden kielten yliopistollisiin oppiaineisiin kirjallisuuden tutkimus kuitenkin edelleen Suomessa kuuluu.

Perinteisissä oppiaineiden ja oppituolien nimissä filologia siis elää. Eri asia on, kuinka paljon näiden filologia-aineiden sisällä on tosiasiassa tutkijoita ja opettajia, jotka kuvaisivat omaa työtään filologiseksi. Jo vuonna 2000 Fred Karlsson ja Nils Erik Enkvist totesivat seuraavaa: ”1990-luvun Suomesta saa melkein etsiä vieraiden kielten filologeja sanan ahtaimmassa ja perinteisimmässä merkityksessä, siis vanhojen tekstien tulkitsijoita”. Sittemmin (2011) Harry Lönnroth on kommentoinut samaa tilannetta: ”useimmissa suomalaisissa yliopistoissa filologia tieteenä ja professiona on enää muisto vain, vaikka moni kielten opiskelija saattaakin vielä kutsua itseään filologiksi erottautuakseen esimerkiksi käännöstieteen opiskelijoista”. *Filologia*-sanan käyttäminen käännöstieteistä erottumisen keinona on yksi tapa ottaa kantaa *filologia*-sanan merkitykseen – joskaan oppiaineiden perinteisissä nimissä, joissa on usein korostunut kääntämisen ja tulkkauksen alasta erottautuminen, filologia ei kääntämisen vastakohtana ole tarkoittanut pelkästään kielentutkimusta vaan laajasti ottaen kielen ja kirjallisuuden tutkimusta.

Filologian historian avulla voidaan ymmärtää, kuinka on tultu tulokseen, jossa sanalle voidaan käytännössä määritellä useita merkityksiä, jotka kuitenkin myös risteävät keskenään. Kuva muuttuu lisäksi huomattavasti, kun tarkastelemme eri kielialueiden tapoja käyttää sanaa: esimerkiksi ranskassa *philologie* ja saksassa *Philologie* ovat viitanneet tekstien editointiin ja tulkintaan, kun taas englanninkielisillä alueilla *philology* on painottunut pikemminkin historiallisten lähteiden kielen tutkimukseen (ks. esim. Adamson & Aures-Bennet 2011). Painolastia on siis melko tavalla ja väärinymmärrysten riski olemassa. Historiallisen taustan hahmottaminen auttaa kuitenkin ymmärtämään, miksi filologiaa ei aina ole helppoa rajata ja määritellä: koska se on toiminut kasvualustana monenlaisille humanistisen tutkimuksen osa-alueille, sen tieteenalaidentiteettiä ovat usein muovanneet vastakkainasettelut – näin on tietysti monilla muillakin aloilla.

Mitä tulee filologian harjoittamiseen nykyään, tilanne ei kuitenkaan ole ehkä aivan niin synkkä kuin edellä siteeratut tutkijat ovat antaneet ymmärtää. Filologista tutkimusta tehdään tavalla tai toisella edelleen

niissä kotimaisissa yliopistoissa, joissa sitä on aiemminkin harjoitettu, joskin oppiainekohtaiset erot ovat varmasti merkittäviä. Viime aikoina moni on myös ollut aistivinaan jonkinlaista filologian renessanssia. Kenties näin onkin: filologian itsereflektio tieteenalana näyttää lisääntyneen. Viime vuosina on ilmestynyt useita artikkelikokoelmia, joissa tarkastelun kohteeksi on nostettu filologian olemus ja relevanssi, esimerkiksi *The Powers of Philology: Dynamics of Textual Scholarship* (Gumbrecht 2003), *Philology and Its Histories* (Gurd 2010), *World Philology* (Pollock, Elman & Chang 2015) ja *Philology Matters! Essays on the Art of Reading Slowly* (Lönnroth 2017). Lisäksi *Maailmankirjallisuuden filologia* on juuri suomennettu (Auerbach 2018 [1967]), ja hiljattain on perustettu uusi aikakauslehti *Philological Encounters*. Nämä ovat esimerkkejä, joissa filologia on sanana nostettu näkyviin, mikä on merkille pantavaa. Kielentutkimuksen alojen historiassa käytyjen itsenäistymistaistelujen jättämät arvet haalistuvat vääjäämättä. Kielitieteen – jonka voinee sanoa tarkastelevan kieltä sen itsensä vuoksi, ei tarkoituksenaan tulkita vanhoja tekstejä – voi katsoa irronneen filologiasta 1800-luvun kuluessa, eikä irtautuminen käynyt aivan kivuttomasti. Vastakkainasettelu filologian ja kielitieteen välillä onkin aiemmin merkinnyt kuohuvia tunteita, koska taustalla on ollut usein ajatuksia tieteiden paremmuusjärjestyksestä. Nykyään suhteiden ei tarvitse enää olla huonot: samat ihmiset voivat katsoa olevansa sekä filologeja että kielitieteilijöitä, ja voi riippua tilanteesta tai kysyjästä, sanooko tutkija olevansa filologi vai historiallisen kielitieteen harjoittaja.

Palataan kysymykseen filologian nykytilasta. Esimerkiksi amerikkalainen sanskritin tutkija Sheldon Pollock on jokseenkin provokatiivisesti syyttänyt (2009, 947) filologeja ilmeisesti jostain, mikä on lähinnä huonoa brändäystä: ”we have failed spectacularly to conceptualize our own disciplinarity”. Samassa yhteydessä Pollock on myös ilmaissut huolensa siitä, jääkö filologia ylipäätään eloon: hän havaitsee yleisesti ihmistieteiden kentällä mm. teorian liikakasvua (engl. *hypertrophy of theory*), mikä on johtanut havainnoitavan materiaalin syrjäytymiseen tutkimuksen ytimestä. Muita hänen mainitsemiaan asioita, joiden kehitystä tulee erityisesti seurata, ovat tekstuaalisen suhde oraaliseen ja visuaaliseen (tekstuaalinen uhkaa jäädä jälkimmäisten jalkoihin) sekä vieraiden kielten,

erityisesti historiallisten, osaaminen, jossa on havaittavissa selvää taantumaa. Nykytilaan lienee johtanut perinteistä filologiaa teoreettisemmin orientoituneen kielentutkimuksen kasvu ja suosio erityisesti 1950-luvun lopulta lähtien: teorialähtöinen tulokulma – vaikkapa Noam Chomskyn tapaan – oli pitkän aikaa keskeisempi ja tuoreemman tuntuinen kuin aineistolähtöinen. Oma vaikutuksensa asiaan ainakin nykykielissä on varmaankin ollut myös sillä, että uusimpaan teoriaan tarttuminen on ollut tutkijan omalle uralle hyväksi – toisin kuin perinteisempi filologinen ns. lähiluku, jonka tarkoituksena on avata kieltä ja tekstin tuottanutta kulttuuria analyyttisen, täsmällisen ja perusteellisen tarkastelun avulla. Toisaalta hyödyllinen lähestymistapa filologisessa metakeskustelussa voisi olla niin ikään Pollockin (2009, 934) esittelemä tekstuaalisuuden teoria (engl. *theory of textuality*), jota filologia edustaa siinä, missä lingvistiikka edustaa teoriaa kielestä. Monien filologiksi itsensä mieltävien mielestä luonnehdinta filologisesta lähestymistavasta ”dataorientoituneena”, siinä missä lingvistinen lähestymistapa on ”teoriaorientoitunut” (Hogg 1994, 3), vaikuttaa myös osuvalta, mutta on toki huomattava, että monia kielitieteen osa-alueita – esimerkiksi korpuslingvistiikkaa – leimaa aineistokeskeisyys tai -lähtöisyys, eivätkä tällaista data vs. teoria -kahtiajakoa varmasti monet lingvistit allekirjoittaisikaan.

Perinteisellä filologialla ei ole ollut tarvetta erityisesti irtisanoutua esimerkiksi kirjallisuudentutkimuksesta tai juuri mistään muustakaan; filologi on ollut vapaa kysymään tekstiltään relevanteiksi katsomiaan kysymyksiä ja tarvittaessa hyödyntämään sekä kielentutkimuksen että muiden alojen tutkimuksen menetelmiä ja teorioita. Osa filologeista katsoo tekstiä mahdollisesti vain kielen kannalta, osalle puolestaan tekstin näkeminen sanataiteena on olennaista – tosin esimerkiksi genreä kumpikaan ei voi jättää huomioimatta. Olipa tutkimuskysymys mikä tahansa, filologiseksi tutkimuksen tekee tutkijan asenne tekstin historiassa ilmenemistä kohtaan: filologi ei sivuuta kysymyksiä siitä, mikä on tekstin hahmo ja miten siihen on tultu. Mikäli näitä kysymyksiä ei millään tavalla pidetä esillä, kysymyksessä ei voine olla filologinen tutkimus.

## 2.3. Tekstin ikä ja menneisyyden rajoitukset

Filologit voivat tutkia oikeastaan mitä tahansa ilmiötä, joka on jollain tavalla läsnä ja havainnoitavissa teksteissä. Joitakin pienimpiä yhteisiä nimittäjiä filologiselle tutkimukselle täytynee kuitenkin voida osoittaa. Filologian olemukseen on aina kuulunut vanhojen tekstien tulkitseminen: alun perin kreikan- ja latinankielisten sekä muilla ns. kuolleilla kielillä kirjoitettujen tekstien tulkitseminen, sittemmin muilla, uudemmilla kielillä kirjoitettujen. Voidaan kuitenkin aiheellisesti kysyä, onko aina mukana oltava jonkinlainen historiallinen komponentti, jotta voimme puhua filologiasta. Minkä ikäinen teksti, teos tai fyysinen kirja on historiallinen tässä mielessä eli riittävän kaukaa menneisyydestä ollakseen filologisessa mielessä kiinnostava? Jo vaikkapa kymmenen vuotta vanha teksti on kirjoitettu selkeästi eri ajassa kuin missä me elämme nyt, ja ennen internetaikaa tuotetut aineistot on kirjoitettu hyvin erilaiselle alustalle kuin merkittävä osa viime vuosien teksteistä. Tässä mielessä varsinaista aikarajaa ei oikeastaan ole mahdollista eikä tarpeellistakaan asettaa. Tietenkin on hyvin harvinaista, että suhteellisen uusia tekstejä tutkiva asemoisi itsensä ensisijaisesti filologiksi.

Kielikohtaiset erot ovat niin ikään suuria, mitä tulee tekstien ikään; esimerkiksi vanhimmat kreikankieliset tekstit ovat reilusti yli 2 000 vuotta vanhempia kuin vanhimmat suomenkieliset tekstit. Englannin kieltä on säilynyt 600-luvulta alkaen, vaikka tämä vanhin aineisto onkin säilynyt meidän aikoihimme jonkin verran nuoremmissa käsikirjoituksissa, kuten on myös monissa muissa kielissä. Hieman karrikoiden voisikin esittää kysymyksen siitä, pitääkö tutkimusaineistossa olla jokin tekstuaalinen haaste tai epävarmuustekijä, jotta kyse on filologiasta. Mahdollisesti kyse on siitä, että filologinen luenta näyttää edellyttävän ajallista tai paikallista – sanalla sanoen ehkäpä kulttuurista – etäisyyttä sille tarpeellisen hitauden saavuttamiseksi (ks. luku 2.4.). Koska filologia tutkii menneisyyden tekstejä, suurimman rajoituksen tutkimusaiheiden suhteen muodostaa se, mitä tekstejä meillä ylipäätään on luettavissamme tänä päivänä. Esimerkiksi kielentutkimuksen historian tutkimusta rajoittaa huomattavasti se, että vaikka pystymme antiikin teoksia lukemalla päättelemään, että stoalainen filosofia vaikutti

huomattavasti siihen, millaiseksi kielentutkimus muotoutui, emme välittömien lähteiden puuttuessa pysty erityisen tarkasti sanomaan, millä tavoin se tapahtui.

## 2.4. Uusia tuulia

Tutkimusaiheissa trendit vaihtelevat, ja uusia ilmiöitä nousee pinnalle: esimerkiksi englannin kielen historian tutkimuksessa myöhäisuusenglanti (engl. *Late Modern English*), jota puhuttiin ja kirjoitettiin suunnilleen 1700-luvulta 1900-luvun alkuun, on noussut merkittäväksi uudeksi tutkimuskohteeksi vasta viime vuosina. Tähän on syynä nykyään jo riittäväksi koettu ajallinen etäisyys esimerkiksi 1800-luvun kieleen ja aineistoihin mutta varmasti myös se, että esimerkiksi muinaisenglannin eli noin tuhat vuotta myöhäisuusenglantia vanhemman kielimuodon taitajia ei ole tutkimuksia kirjoittamassa tai lukemassa nykyään lainkaan niin paljon kuin vielä muutama vuosikymmen sitten. Eri asia sitten on, pitävätkö tämän myöhemmän kauden tutkijat itseään Suomessakaan filologeina – ehkä pikemminkin englannin kielen historiaan suuntautuneina lingvisteinä tai jonkin tarkemmin määritellyn kielitieteen osa-alueen tutkijoina.

Filologia siis elää ja kehittyy. Vuodesta 1990 lähtien on puhuttu *uudesta filologiasta* (engl. *New Philology*), sittemmin myös havainnollisemmin *materiaalisesta filologiasta*. Jälkimmäisen nimityksen käyttäminen lienee paitsi selkeämpää (myös erotuksena *uusfilologiasta*, jolla on aiemmin viitattu modernien kielten tutkimukseen) myös realistisempaa: ei ole tarkoitus implikoida, että ”uusi” olisi jollain tavalla tullut ”vanhan” tilalle, eikä ”vanha” myöskään ole vanhentunut; ei ole myöskään kategorisesti totta, että ”uusi” olisi jotakin aivan ennen harjoittamatonta. Mistä sitten on kysymys?

Materiaalista filologiaa luonnehtii paluu käsikirjoituksiin pitkän editiokeskeisen ja toistaiseksi melko lyhyen korpuspainotteisen aikakauden jälkeen. Asian tarkasteleminen on hyvä aloittaa tekemällä ero käsitteiden *teos*, *teksti* ja *artefakti* välille. Teoksen voidaan ymmärtää olevan kirjoittajan intention mukainen sisältö tietyssä kielellisessä muodossa.

Teksti puolestaan on yksi teoksen esiintymistä, siis yksi tekstivariantti, joka ei kuitenkaan ole itse teos eikä voi täysin vastata alkuperäisen teoksen kirjoittajan pyrkimystä, sillä tekstin sisältö ja muoto muuttuvat, kun sitä kopioidaan tai painetaan uudelleen. Artefakti viittaa tekstin materiaaliseen muotoon: käsikirjoitukseen tai painotuotteeseen, jossa teksti on säilynyt. Huomautettakoon kuitenkin, että on tavallista puhua tekstistä laajemmin viitaten ylipäätään jonkin teoksen geneeriseen tekstiin yksittäisen esiintymän sijaan, kuten tehdään myös tässä artikkelissa.

Perinteisen filologisen tarkastelun kohteena on ollut juuri teos, jonka hahmosta todistaneet yksittäiset käsikirjoitukset ovat erikseen tai itsessään olleet kiinnostavia vain tiettyyn pisteeseen saakka. Materiaalinen filologia sen sijaan on kiinnostunut artefaktitasosta: meille säilyneen artefaktin syntymisen sosiaalisesta, kulttuurisesta ja taloudellisesta kontekstista ja sen tekemiseen käytetyistä tekniikoista, käsikirjoituksesta (ja painetusta materiaalista) kokonaisvaltaisesti fyysisenä objektina kaikkine siihen liittyvine elementteineen (nimiölehdet, prologit ja epilogit, marginaalit) sekä tekstin ja artefaktin välisestä suhteesta (tekstin asettelu sivulle, kirjaintyyppi, mahdollinen kuvitus tai otsikointi). Kiinnostavaa ei ole (ainoastaan tai ensisijaisesti) se, mikä on kunkin tekstin alkuperäinen kielellinen hahmo, vaan kunkin tekstivariantin oma yksilöllinen tarina ja sen synty- ja käyttökonteksti. Näin keskeiseksi tutkimuskohteeksi nousee alkuperäisen sisällön ja kieliasun sijaan tekstin välittyminen tuottajalta kuluttajalle ja sen käyttö. Tekstuaalinen varianssi nousee tekstuaalisen auktoriteetin tilalle tai ainakin rinnalle. Laajemmin ajateltuna fokuksessa ovat teoksen tuottaneen kulttuurin sijaan tekstin ja artefaktin tuottanut kulttuuri ja kulttuuriset implikaatiot.

## 3. Miten filologista tutkimusta tehdään

Filologian olemusta hahmoteltaessa on ollut suosittua todeta, että sen ytimessä on hidas lukeminen. Ajatus tulee Friedrich Nietzscheltä (1844–1900), joka muun muassa oli kreikkalaisen filologian professori Baselin

yliopistossa, ja edelleen sitä teki tunnetuksi venäläis-amerikkalainen kielen- ja kirjallisuudentutkija Roman Jakobson (1896–1982). Tähän hitaan lukemisen prosessiin on perinteisesti filologin hyveiksi liitetty tietynlainen tulkinnan hienovaraisuus, täsmällisyys ja herkkyys yksityiskohdille sekä näiden edellä mainittujen toteutumisen edellytyksenä mitä ilmeisimmin kärsivällisyys. Nämä eivät ehkä ole tyydyttäviä ohjeita esimerkiksi tutkielmansa metodologiaan liittyvää ohjausta kaipaavalle opiskelijalle, mutta ne ovat yhtä kaikki realistisia huomioita tutkimuksen tekemisen luonteesta. Tässä luvussa tarkastelemme lähemmin filologisen tutkimuksen metodologiaa. Lähestymme aihetta useista eri näkökulmista: funktionaalisesti tieteenalan sisältä, sen lähialojen kannalta sekä tutkimuskysymysten ja ammatti-identiteettien kannalta.

## 3.1. Monimuotoinen lähestymistapa

Filologiasta on luontevaa puhua tieteellisenä lähestymistapana tai näkökulmana; filologisilla metodeilla puolestaan täsmällisesti ottaen ymmärretään ennen kaikkea kriittisten tekstieditioiden valmistamisen periaatteita. Kyse on metodeista, joita filologin on tarkoituksenmukaista ymmärtää sitä paremmin, mitä vanhempia ja käytetympiä tekstejä hän tutkii, vaikka ei itse päätyisikään tekemään tieteellistä editiota. Filologinen orientaatio on periaatteessa hyvin käytännöllinen: pyrkimyksenä on tekstin ymmärtäminen mahdollisimman hyvin. Erilaisia teoreettisia viitekehyksiä on mahdollista hyödyntää sen mukaan, mikä on tarpeellista. Filologit ovat toimineet näin esimerkiksi pragmatiikan, sukupuolentutkimuksen, kielihistorian ja kirjallisuusteorioiden kanssa. Wegmann (2014, 27) puhuukin mitä moninaisimpia yhdistelmiä suvaitsevasta polymorfisuudesta suorastaan filologian tavaramerkkinä. Toisinaan ei ole edes erityisen selvää, missä teoreettisten viitekehysten tai lähestymistapojen rajat kulkevat. Esimerkiksi filologian ja uuden lähialan läheisyydestä käy historiallinen sosiolingvistiikka, jonka kenttää *Journal of Historical Sociolinguistics* kuvailee verkkosivuillaan seuraavasti: ”aspects of language and society in the past including (but not limited to) the social embedding of language variation and change, issues of

language contact and conflict, historical multilingualism, the social stratification of writing skills, the development of language norms and the impact of language ideologies". Näissä fokuksissa ei ole välttämättä sinänsä mitään vierasta sellaisillekaan filologeille, jotka eivät erityisesti identifioidu sosiolingvisteiksi. Provokatiivisesti voisikin kysyä, voiko tällaista tutkimusta tehdä ilman filologista ymmärrystä tekstien ja kontekstien luonteesta. Sosiolingvistiikan ohella merkityksiä konteksteissaan tutkivaa pragmatiikkaa voidaan yhdistää filologian kanssa hyödylliseksi lähestymistavaksi (Mäkilähde 2019).

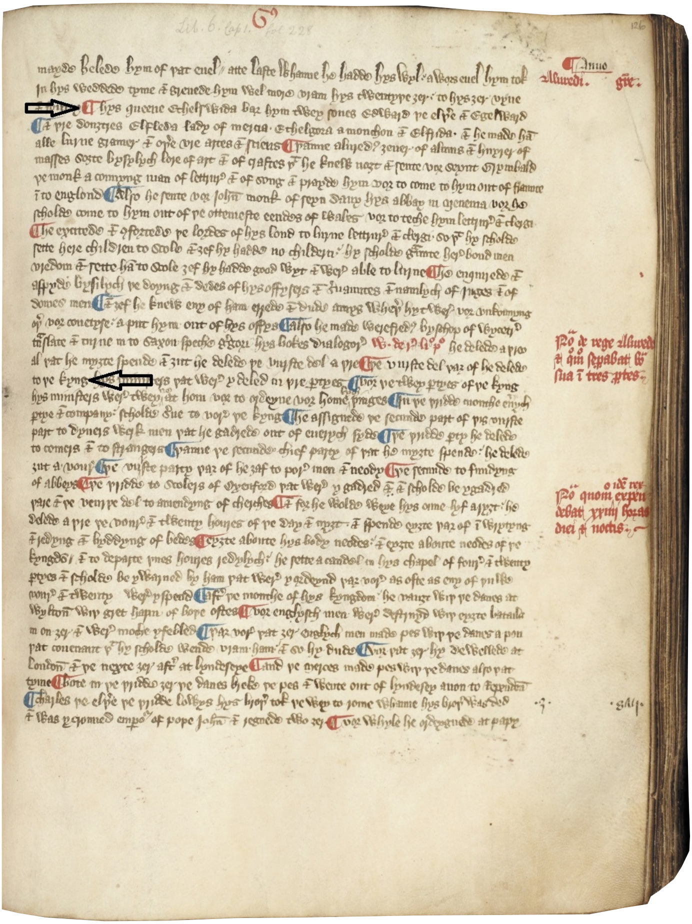
Esimerkkinä materiaalisen filologian yhdistämisestä kielitieteeseen, erityisesti pragmatiikkaan, voidaan tarkastella vaikkapa hiljattain esitettyä neliportaista menetelmää, jota voi soveltaa monenlaisiin samaa teosta heijasteleviin teksteihin. Menetelmän kehittäjillä (Carroll, Peikola, Salmi, Varila, Skaffari & Hiltunen 2013) aineistona on *Polychronicon*, suosittu keskiaikainen latinankielinen maailmanhistoria, joka on säilynyt Euroopassa yli sadassa käsikirjoituksessa. Englanniksi historian käänsi John Trevisa 1300-luvun lopulla. Menetelmässä lähdetään liikkeelle yhden teoksen eri tekstivarianttien (tekstien) tunnistamisesta editioiden ja muun lähdemateriaalin avulla; esimerkiksi Trevisan *Polychronicon*-käännöksiä eli tekstivariantteja on olemassa neljätoista. Sitten edetään kunkin tekstin materiaaliseen kuvaukseen, jossa tarvitaan alkuperäisiä tai digitoituja käsikirjoituksia sekä muun muassa paleografian eli vanhojen käsialojen tutkimuksen taitoja. Seuraavaksi eri tekstien materiaalisia kuvauksia vertaillaan keskenään: millaisille alustoille eri tekstit on kirjoitettu, mitä kirjaintyyppiä eri kirjurit ovat käyttäneet, miten teksti on jaettu riveille ja sivuille, onko käsikirjoituksessa kuvitusta, miten tekstin eri osat on nimetty tai numeroitu jne. Viimeisessä vaiheessa tekstien kielellisiä ja visuaalis-materiaalisia eroja selitetään pragmatiikan käsitteiden avulla tekstin (engl. *textual*), vuorovaikutuksen (engl. *interactional*) ja asenteen (engl. *stance*) tasolla, mikä tuottaa tarkemman analyysin aineistosta kuin pelkästään kieleen keskittyvä pragmatiikka tai materiaaliorientoituneet kodikologian[2] tai paleografian alat yksinään. Yleisemmin materiaalinen filologia antaa kielentutkijalle tarpeellista kontekstuaalista tietoa erityisesti silloin, kun tekstin tuottaja ja vastaanottaja ovat jääneet anonyymeiksi tai heistä ei ole muuta tietoa

kuin nimi. Esimerkiksi kirjan koko, tekstin asettelu sen lehdille sekä se, mitä tekstejä siihen on sidottu ja miten käsikirjoitus on kuvitettu tai koristeltu (tai ei ole), kertovat (fyysisen) kirjan käyttötarkoituksesta, tekijästä ja lukijasta, jotka kaikki heijastuvat myös sen kieliasuun.

## 3.2. Tekstikritiikki ja editiotyypit

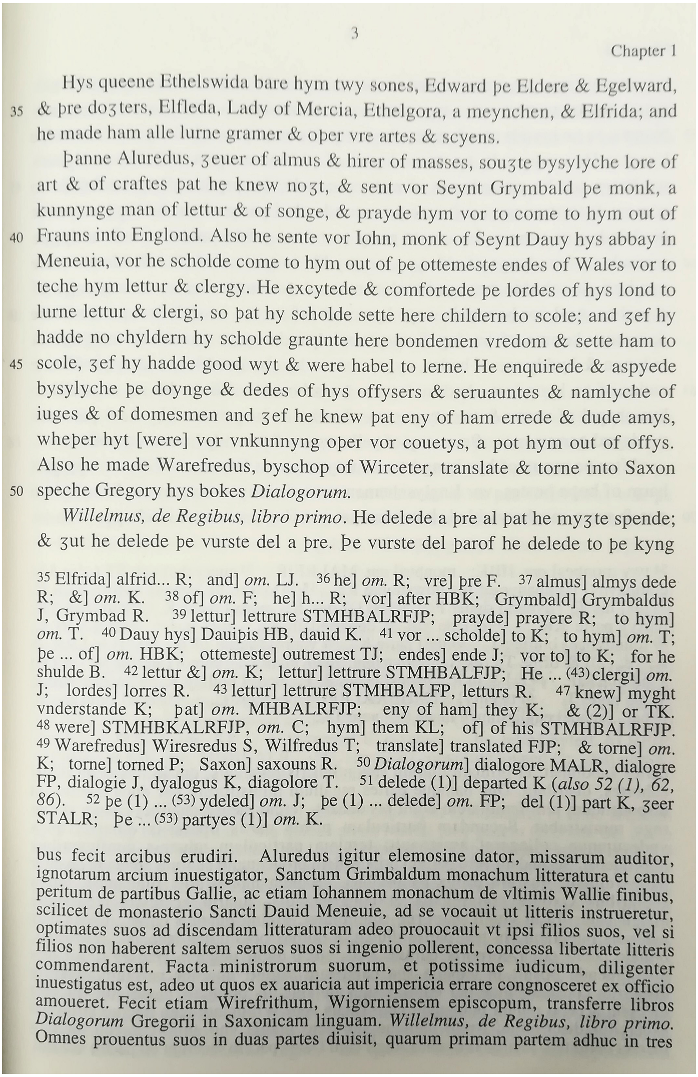
Kuten totesimme, filologisen tutkimuksen pyrkimyksenä on tekstin mahdollisimman syvällinen ymmärtäminen. Tätä ymmärrystä voidaan sitten soveltaa muihin tutkimuksen tavoitteisiin, jotka voivat liittyä kielen-, kirjallisuuden- tai kulttuurintutkimukseen. Tekstin ymmärrystä symboloi ja myös käytännössä edustaa kriittinen editio. Se on teksti, jonka tulee perustua olemassa olevien käsikirjoitusten perinpohjaiseen tuntemukseen. Kriittinen editio muodostuu tietynlaisen prosessin, tekstikritiikin, tuloksena. Tällä tarkoitetaan tieteellistä menetelmää, joka pyrkii selvittämään tekstin alkuperäisimmän mahdollisen muodon vertailemalla siitä säilyneitä – mahdollisesti kovinkin erilaisia – versioita ja rekonstruoimalla niiden perusteella sellaisen tekstimuodon, jonka voi perustellusti arvioida vastaavan alkuperäisen kirjoittajan intentiota. Kriittinen editio sisältää tekstin lisäksi ns. kriittisen apparaatin (lat. *apparatus criticus*) eli varsinaiseen tekstiin liittyvän erillisen osion, jossa erilaiset tekstin lukutavat ja käsikirjoitustraditioon liittyvät merkinnät sijaitsevat. Kriittisessä editiossa sekä teksti että kriittinen apparaatti perustuvat relevanttien käsikirjoitusten tai painosten ja mahdollisten edeltävien kriittisten editioiden konsultoimiseen.

Siinä missä kriittinen editio voi olla erinomainen lähde historioitsijalle tai muulle ennen kaikkea teoksen sisällöstä kiinnostuneelle, se ei auta vaikkapa morfologista vaihtelua tutkivaa, joka tarvitsee mahdollisimman autenttisen aineiston, jota ediittori ei ole muuttanut. Tällöin nousee tarpeelliseksi diplomaattinen editio, jolla tarkoitetaan sellaista tieteellistä editiota, jonka tarkoituksena on esittää käsikirjoitus tai painettu teksti mahdollisimman uskollisesti alkuperäiselle fyysiselle tekstille. Tällöin ediittori ei pyri yhdenmukaistamaan kielellisiä muotoja tai kirjoittajan käytänteitä (kuten välimerkkejä) vaan välittämään luotettavan kuvan



Kuva 1. *Polychronicon*-käsikirjoituksen sivu: Manchester, Chetham's Library MS Mun A.6.90, f. 126r. Kuvaan lisättyjen nuolien välissä on edition tekstikohta (kuva 2). Julkaistu kirjaston luvalla: Chetham's Library, Manchester.

yhdestä tekstistä. Sen tarkoitus ei siis ole rekonstruoida taustalla häämöttävää teosta vaan dokumentoida yhtä tekstivarianttia. Nykyään tutkijan aineistona voi olla myös digitaalisia kuvia, jotka kirjaimellisesti antavat vielä tarkemman kuvan tutkittavasta materiaalista kuin editio tai faksimile[3]. Sivu sivulta valokuvattu käsikirjoitus, joka on nähtävillä



3

Chapter 1

Hys queene Ethelswida bare hym twy sones, Edward þe Eldere & Egelward, & þre doȝters, Elfleda, Lady of Mercia, Ethelgora, a meynchen, & Elfrida; and he made ham alle lurne gramer & oþer vre artes & scyens.

Þanne Aluredus, ȝeuer of almus & hirer of masses, souȝte bysylyche lore of art & of craftes þat he knew noȝt, & sent vor Seynt Grymbald þe monk, a kunnynge man of lettur & of songe, & prayde hym vor to come to hym out of Frauns into Englond. Also he sente vor Iohn, monk of Seynt Dauy hys abbay in Meneuia, vor he scholde come to hym out of þe ottemeste endes of Wales vor to teche hym lettur & clergy. He excytede & comfortede þe lordes of hys lond to lurne lettur & clergi, so þat hy scholde sette here childern to scole; and ȝef hy hadde no chyldern hy scholde graunte here bondemen vredom & sette ham to scole, ȝef hy hadde good wyt & were habel to lerne. He enquirede & aspyede bysylyche þe doynge & dedes of hys offysers & seruauntes & namlyche of iuges & of domesmen and ȝef he knew þat eny of ham errede & dude amys, wheþer hyt [were] vor vnkunnyng oþer vor couetys, a pot hym out of offys. Also he made Warefredus, byschop of Wirceter, translate & torne into Saxon speche Gregory hys bokes *Dialogorum*.

*Willelmus, de Regibus, libro primo*. He delede a þre al þat he myȝte spende; & ȝut he delede þe vurste del a þre. Þe vurste del þarof he delede to þe kyng

35 Elfrida] alfrid... R; and] *om*. LJ. 36 he] *om*. R; vre] þre F. 37 almus] almys dede R; &] *om*. K. 38 of] *om*. F; he] h... R; vor] after HBK; Grymbald] Grymbaldus J, Grymbad R. 39 lettur] lettrure STMHBALRFJP; prayde] prayere R; to hym] *om*. T. 40 Dauy hys] Dauiþis HB, dauid K. 41 vor ... scholde] to K; to hym] *om*. T; þe ... of] *om*. HBK; ottemeste] outremest TJ; endes] ende J; vor to] to K; for he shulde B. 42 lettur &] *om*. K; lettur] lettrure STMHBALFJP; He ... (43) clergi] *om*. J; lordes] lorres R. 43 lettur] lettrure STMHBALFP, letturs R. 47 knew] myght vnderstande K; þat] *om*. MHBALRFJP; eny of ham] they K; & (2)] or TK. 48 were] STMHBKALRFJP, *om*. C; hym] them KL; of] of his STMHBALRFJP. 49 Warefredus] Wiresredus S, Wilfredus T; translate] translated FJP; & torne] *om*. K; torne] torned P; Saxon] saxouns R. 50 *Dialogorum*] dialogore MALR, dialogre FP, dialogie J, dyalogus K, diagolore T. 51 delede (1)] departed K (*also 52 (1), 62, 86*). 52 þe (1) ... (53) ydeled] *om*. J; þe (1) ... delede] *om*. FP; del (1)] part K, ȝeer STALR; Þe ... (53) partyes (1)] *om*. K.

bus fecit arcibus erudiri. Aluredus igitur elemosine dator, missarum auditor, ignotarum arcium inuestigator, Sanctum Grimbaldum monachum litteratura et cantu peritum de partibus Gallie, ac etiam Iohannem monachum de vltimis Wallie finibus, scilicet de monasterio Sancti Dauid Meneuie, ad se vocauit ut litteris instrueretur, optimates suos ad discendam litteraturam adeo prouocauit vt ipsi filios suos, vel si filios non haberent saltem seruos suos si ingenio pollerent, concessa libertate litteris commendarent. Facta ministrorum suorum, et potissime iudicum, diligenter inuestigatus est, adeo ut quos ex auaricia aut impericia errare congnosceret ex officio amoueret. Fecit etiam Wirefrithum, Wigorniensem episcopum, transferre libros *Dialogorum* Gregorii in Saxonicam linguam. *Willelmus, de Regibus, libro primo*. Omnes prouentus suos in duas partes diuisit, quarum primam partem adhuc in tres

Kuva 2. *Polychronicon*-edition sivu. Lontoon British Libraryssä olevan käsikirjoituksen (Cotton Tiberius D.VII) teksti on sivun yläosassa. Keskellä näkyy kriittinen apparaatti, jossa mainitaan muissa käsikirjoituksissa esiintyvät variantit; Manchesterin käsikirjoituksen (kuva 1) lyhenne on M. Alimpana on latinankielinen lähtöteksti. Julkaistu kustantajan luvalla: Ronald Waldron (toim.). 2004. *John Trevisa's Translation of the* Polychronicon *of Ranulph Higden, Book VI*. [Middle English Texts, vol. 35]. Heidelberg: Universitätsverlag Winter, 3.

esimerkiksi jonkin kirjaston verkkosivuilla, ei kuitenkaan itsessään ole editio vaan kokoelma digitaalisia kuvia ilman lähdekirjallisuutta ja filologin panosta. Editiot puolestaan ovat sekä toimittajansa tutkimuksen tuloksia että tutkimusaineistoja käyttäjilleen.

## 3.3. Kuka on filologi?

Tutkimuksen tekemisen tapaa voi tarkastella myös ammatti-identiteettien kautta. Filologisen koulutuksen saaneiden ammatti-identiteetti vaihtelee: esimerkiksi klassillisen filologian parissa osa mieltää itsensä kielitieteilijöiksi, osa identifioituu yleisesti antiikintutkijoiksi ja eräille tarkempi erikoistumisala, vaikkapa papyrologia, on ensisijainen ammatti-identiteetin lähde. Nykykielten filologit ovat – ainakin Suomessa – usein koulutukseltaan pikemminkin kielitieteilijöitä kuin kirjallisuuden- tai kulttuurintutkijoita, vaikka filologia-aineiden opetussuunnitelmiin tyypillisesti kuuluukin myös kirjallisuusopintoja ja kohdekielen kannalta relevantteihin kulttuureihin ja historiaan tutustumista. Moni filologi varmasti identifioituu tietyn kielen tutkijaksi, ei niinkään yleislingvistiksi. Myös materiaalisen filologian tutkijalla saattaa olla lingvistinen koulutus, mutta hänen työnsä keskiössä voi silti olla tekstin kielellisten piirteiden sijaan vaikkapa kirjoitusalusta, käsiala tai kirjaintyyppi. Hän toimii siis myös kodikologian ja paleografian aloilla tai käyttää niiden menetelmiä. Materiaalinen filologi voi siis huomioida tutkimansa tekstin ohella muut samaan käsikirjoituskirjaan sidotut tekstit ja niiden yhteydet varsinaiseen tutkimuskohteeseen tai tarkastella esimerkiksi taivutusmorfologisen vaihtelun sijaan tai lisäksi kirjainmerkkien muodoissa esiintyvää variaatiota. Kodikologiasta, paleografiasta ja tekstikritiikistä voidaan puhua ns. tekstuaalitieteinä (engl. *textual scholarship*) samoin kuin muista tekstien kirjoittamista, variaatiota ja välittymistä tutkivista tieteistä, joiden intressit tulevat usein varsin lähelle filologisia intressejä.

Metodologisesti pelkkä filologia ei välttämättä kuvaakaan erityisen tarkasti monenkaan tutkijan tekemisen tapaa, minkä vuoksi alojen nimitykset ja ammatti-identiteetit ovat jatkuvan keskustelun ja uudelleenmäärittelyn kohteena: kuten Wegmann (2014, 27) toteaa, tieteenalan

itseymmärryksestä on tullut pluralistista. Yleisesti identiteettien suhteen täytyy todeta myös, että lukkiutuminen ja raja-aitojen rakentelu eivät ole tätä päivää; tutkijan on mahdollista ja tarpeellistakin omaksua uransa aikana monia erilaisia identiteettejä. Monille kielentutkimuksen ammattilaisille keskeisin identiteetti on, kuten sanottu, kuitenkin sidoksissa erikoistumiskieleen, vaikka esimerkiksi germanisteista ja romanisteista kuuleekin syystä tai toisesta puhuttavan yhä vähemmän.

Filologisella otteella tutkimusta tekevän pitää luonnollisesti tuntea tutkimansa kielimuoto hyvin. Nykykielten oppiaineissa on vaihtelevasti tarjolla kielten varhaisvaiheiden opetusta, mutta peruskurssien jälkeen kielimuotoa voi opiskella myös itsenäisesti, ja sanaston ja kieliopin tuntemus joka tapauksessa karttuu tutkimusta tehdessä. Kielitaitoon liittyy myös oman alan perustyökalujen eli (historiallisten) sanakirjojen ja kielioppien käyttö. Latinan ja muinaiskreikan osalta voidaan sanoa, että opetus yleensä keskittyy klassisimman vaiheen kielimuotoon; ajalliset, paikalliset ja genreen liittyvät erityispiirteet muodostavat haasteen näissäkin kielissä. Tekstien ”hidas lukeminen” edellyttää kielimuodon hallinnan lisäksi myös tutkittavan kauden historian ja kulttuurin tuntemusta, jotta anakronismeilta (eli karkeimmillaan ilmiöiden sijoittamiselta väärään historialliseen aikakauteen) voidaan välttyä. Tekstien kirjoittajiin ja vastaanottajiin on tarpeellista tutustua mahdollisimman hyvin, samoin on tunnettava heidän kulttuurinen viitekehyksensä, joka vaikuttaa muun muassa intertekstuaalisuuteen eli siihen, miten eri tekstien välille rakennetaan yhteyksiä esimerkiksi sitaateilla ja alluusioilla. Filologia on kontekstualisoivaa lukemista, eikä kontekstiin tutustuminen käy hetkessä vaan edellyttää usein pidempää paneutumista opintojen aikana. Lisäksi on mainittava, että yksittäisen tekstin kontekstiin eivät kuulu ainoastaan siihen liittyvät meille säilyneet tekstit vaan tyypillisesti myös suuri joukko jossain kohtaa traditiota vaikuttaneita tekstejä, jotka ovat peruuttamattomasti kadonneet. Tämä kaikki on sellaista, mikä filologin on kyettävä jollain tavalla ottamaan huomioon. Tiivistäen voidaan sanoa, että filologinen havainnointi perustuu mahdollisimman syvälliseen ymmärrykseen siitä, millaisissa oloissa käsiteltävänä oleva teksti on syntynyt.

## 3.4. Filologin materiaali

Käytännössä filologista tutkimusta tehdään sekä suoraan alkuperäisaineistoon perustuen (esim. papyrusmateriaali, piirtokirjoitukset, eri-ikäiset käsikirjoitukset, painettu materiaali) että kriittiseen (tai diplomaattiseen) editioon perustuen. Riippuu suuresti aikakaudesta ja kulttuuripiiristä, missä määrin, millaisia ja minkä tyyppisiä editioita on saatavana. Voidaan sanoa, että esimerkiksi klassillisen filologian piirissä homeerisia tekstejä on editoitu kunkin ajan parhaiden tieteellisten käytäntöjen mukaan jo 200-luvun alkupuolelta eaa. Keskiaikaisten käsikirjoitusten välityksellä meille periytyneet antiikin ajan tekstit onkin jo kaikki ainakin kertaalleen editoitu, mutta esimerkiksi bysanttilaisten *skholia*-tekstien (eli varsinaista, tyypillisesti antiikista peräisin olevaa tekstiä kommentoivien reunatekstien) ensieditioita ilmestyy edelleen säännöllisesti. Ottaaksemme esimerkin kotimaasta voimme todeta, että Aleksis Kiven teosten kriittisten editioiden julkaiseminen alkoi vuonna 2010 ja on tätä kirjoitettaessa vielä kesken.

Filologiselle tutkimukselle on ollut leimallista, että tutkimustradition alkuvaiheissa tärkeäksi on noussut erityisesti kaunokirjallisuuden ja ylipäätään kansallisromanttisten ideologioiden valossa merkittävien tekstien tutkimus. Myöhemmin on tullut tilaa myös arkisempien tekstilajien, kuten kirjeiden ja erilaisten asiakirjojen, tutkimukselle. Esimerkiksi voi nostaa vaikkapa Raija Sarasti-Wileniuksen edition *Dear Brother, Gracious Maecenas: Latin Letters of the Gyldenstolpe Brothers (1661–1668)*. Mitä tulee ruotsinkielisiin teksteihin, suuri osa keskiaikaisesta materiaalista on jo kertaalleen editoitu; viimeaikaisista editioista voidaankin mainita Evert Meleforsin *Senmedeltida kopparutvinning i Åtvidaberg speglad i räkenskaper från 1500-talets början* (2013), joka edustaa juuri sellaista materiaalia (tilikirjoja), jonka editointia ei ole koettu ensisijaiseksi, sekä uudelleen editoitavat Pyhän Birgitan ilmestykset (toim. Roger Andersson, vuodesta 2014 lähtien). Editoinnissa on nykyään myös vahva digitaalinen ulottuvuus, mitä kuvastaa esimerkiksi Ville Marttilan väitöskirja: *Creating Digital Editions for Corpus Linguistics: The case of Potage Dyvers, a family of six Middle English recipe collections* (2014) on korpuslingvistisesti suuntautunut dokumentaarinen editio keskiajan lopun kansankielisistä ruokaresepteistä.

Filologista lähestymistapaa voi pitää soveltuvana kaikkeen menneisyydestä meille säilyneeseen tekstiin. Koska on kysymys tekstistä ja menneisyydestä – vaikka ei välttämättä kaukaisestakaan –, voidaan sanoa, että aineisto määrää filologin mahdollisuuksia tutkia enemmän kuin monen muun lähestymistavan edustajan. Aivan kaikkien ilmiöiden tutkiminen ei ole mahdollista kaikkien aikakausien osalta lähdemateriaalin puuttumisen vuoksi. Klassillisen filologian kohdalla aineisto on ollut viimeisen vuosisadan ajan suhteellisen vakiintunutta. Jonkin verran saadaan toki vuosittain tutkittavaksi ennennäkemätöntä kirjallista materiaalia suoraan antiikista, siis ilman keskiaikaisia välikäsiä eli käsikirjoituksia. Esimerkiksi kesällä 2017 raportoitiin Pohjois-Englannista jälleen uusista tekstilöydöistä, ns. Vindolandan tableteista eli kirjoitustauluista, jotka ovat peräisin ensimmäiseltä vuosisadalta jälkeen ajanlaskun alun. Myös yksittäisten auktorien eli kirjoittajien korpus elää: vaikka esimerkiksi antiikista peräisin olevia uusia kaunokirjallisia tekstikokonaisuuksia tuskin enää löytyy, voi attribuutioihin[4] tulla tarkennuksia mm. tietokoneilla toteutettavan laajoihin korpusanalyyseihin perustuvan tutkimuksen avulla. Edes Shakespearen tuotanto ei ole muuttumaton: esimerkiksi vuonna 2016 vahvistettiin Henrik VI -näytelmät William Shakespearen ja Christopher Marlowen yhdessä kirjoittamiksi.

## 3.5. Digitalisaatio ja aineistot

Filologien tutkimusmahdollisuudet paranevat ja helpottuvat suuresti digitaalisten ympäristöjen kehittymisen myötä, kun käyttöön tulee laajoja korpuksia, digitoituja aineistoja ja kaikkien hyödynnettävissä olevia tietokantoja. Tutkimusaineistojen kattavuus tietenkin vaihtelee kielittäin, samoin maksullisten korpusten ja korpusohjelmien saatavuus yliopistoittain. Klassillisen filologian parissa on perinteisesti oltu digitalisaation suhteen hyvin ajan hermolla. Esimerkiksi Thesaurus Linguae Graecae (TLG) on jo vuonna 1972 perustettu projekti, joka nykymuotoisena käsittää valtaosan kaikista kreikankielisistä teksteistä homeerisista teksteistä vuoteen 1453 saakka. Historiansa aikana TLG on ehtinyt olla saatavana magneettinauhoilla, CD-romilla ja vuodesta 2001 verkossa.

Painetun kirjan varhaisesta historiasta kiinnostuneelle anglistille puolestaan Early English Books Online (EEBO) on äärimmäisen arvokas aineistolähde. Se sisältää kuvia kutakuinkin kaikkien englanniksi painettujen kirjojen sivuista uuden ajan alusta (1473–1700), ja siinä on materiaalia myös muilla kielillä. Kielellisten ilmiöiden laaja-alaisempi tutkiminen yli 17 miljoonan kuvan kokoelmasta on kuitenkin vaikeaa, jos haut voivat kohdistua vain nimekkeeseen, kirjoittajaan, asiasanoihin, painovuoteen tai muuhun vastaavaan. Niinpä tätä aineistoa täydentää EEBO-TCP (Text Creation Partnership), joka tuottaa EEBOn kirjojen transkriptioita muun muassa kielentutkijoiden (korpus)aineistoksi sopivina xml-dokumentteina.

Ylipäätään voidaan sanoa, että korpuslingvistiikan kehittymisen myötä yleistyneet kvantitatiiviset eli määrälliset menetelmät heijastuvat myös filologiseen tutkimukseen. Pienten aineistojen hidas lukeminen saattaa vaikuttaa jäävän suurten aineistojen nopean, tietokoneavusteisen määrällisen analyysin jalkoihin, vaikka kontekstiin paneutuminen olisi tarpeen monessa tutkimuksessa – mitä suurempi korpus, sitä vähemmän on tietoa yksittäisten tekstien taustoista tai kielenkäyttökonteksteista.

Kvalitatiivisella eli laadullisella sekä pienempien aineistojen tutkimuksella on edelleen paikkansa, ja siinäkin digitalisaatio auttaa tutkijaa. Luvussa 3.1. mainittua *Polychronicon*-tutkimusta ei olisi voinut toteuttaa eikä siinä esiteltyä menetelmää testata, ellei eri käsikirjoitusten sivuista olisi ollut saatavilla digitaalisia kuvia, sillä tutkitut käsikirjoitukset sijaitsivat seitsemässä eri kaupungissa Britanniassa ja Yhdysvalloissa. Alkuperäisaineiston tutkiminen paikan päällä on materiaaliselle filologille ihanteellinen tilanne, mutta se on usein niin hankalaa tai kallista, että riittävän tarkka kuva-aineisto on käytännöllisempi valinta tutkimuksen perustaksi. Digitoidut aineistot voivat sekä mahdollistaa muuten vaikeasti toteutettavan tutkimuksen että sulkea pois sellaisia tutkimuskysymyksiä, joiden vastaamiseen digitaalinen editio tai kuvamateriaali ei riitä.

# 4. Yhteenveto

Olemme siteeranneet yllä useita tuoreita julkaisuja sekä Suomesta että ulkomailta. Jo näiden kirjojen ja artikkelien olemassaolo osoittaa, että filologia ei ole tieteellisenä lähestymistapana häviämässä. Voitaneen puhua pikemminkin vahvistumisesta – tai sitten on alettu tunnistaa uudempien menetelmien juuret, niiden velka perinteiselle tekstintuntemukselle. Yksittäisten tekstien tai tekstistöjen tutkimusta ja syvällistä tuntemusta tarvitaan vastapainoksi tai täydentäjäksi usein kvantitatiiviselle korpustutkimukselle.

Filologisen tutkimuksen yhteenniputtaminen ei ole aivan yksinkertaista. Tutkimuskohdetta tai -kysymyksiä ei näytä olevan tarpeellista nostaa esille tutkimuksen yhteisiä nimittäjiä haettaessa; filologi säilyttäköön vapautensa kiinnostua kaikentyyppisistä ilmiöistä, joihin tekstin tutkiminen voi tuoda valaistusta. Monelle tutkijalle Suomessa nämä ovat epäilemättä kielen ilmiöitä: menneisyyteen keskittyvää synkronista kielentutkimusta voidaan tehdä filologisella otteella.

Mikä sitten on filologiseksi identifioituvan tutkimuksen kannalta tärkeää? Esiin nousevat ensisijaisesti aineistolähtöisyys ja menetelmät. Kontekstien tuntemus ja huomioonottaminen – se, että tekstin mitään osaa ei lueta irrallisena – merkitsevät tutkimuksen lopputuloksen kannalta sitä, että tutkijan tulkinta ei muodostu anakronistiseksi. Riippumatta tutkimusaineistosta ja sen kielestä filologisen tutkimuksen tehtävänä on yhä uusien sukupolvien voimin tarkastella menneisyyden tekstejä ja kysyä niiltä omasta ajasta kumpuavia relevantteja kysymyksiä.

**Aiheesta lisää:**

Driscoll, Matthew James & Pierazzo, Elena (toim.). 2016. *Digital Scholarly Editing: Theories and Practices*. Open Book Publishers.

Fraistat, Neil & Flanders, Julia. 2013. *The Cambridge Companion to Textual Scholarship*. Cambridge: Cambridge University Press.

Tarrant, Richard. 2016. *Texts, Editors, and Readers. Methods and Problems in Latin Textual Criticism*. Cambridge: Cambridge University Press.

## VIITTEET

1 Kreikkalaisten harjoittamaa filologiaa vielä jonkin verran varhaisempaa oli Intiassa harjoitettu pyhien tekstien tulkinta. Ns. *prōtos heurētēs* ('ensimmäinen keksijä') -kisan voittajaa ei ole ehkä kuitenkaan mahdollista ratkaista kovin täsmällisesti.
2 Kodikologia on laveimmassa merkityksessään koodeksien eli lähinnä keskiajalta ja uuden ajan alusta säilyneiden fyysisten kirjojen sekä niiden historian tutkimusta.
3 Faksimile on painotuotteen tai käsikirjoituksen valokuvajäljennös. Painetut faksimilet ovat jo jääneet digitoitujen aineistojen varjoon.
4 *Attribuutiolla* tarkoitetaan teoksen määrittämistä jonkun tietyn auktorin laatimaksi.

## KIRJALLISUUS

Adamson, Sylvia & Aures-Bennet, Wendy. 2011. Linguistics and philology in the twenty-first century: introduction. *Transactions of the Philological Society* 109:3, 201–206.
Altschul, Nadia. 2010. What is philology? Cultural studies and ecdotics. Julkaisussa: Gurd, Sean (toim.) *Philology and Its Histories*. Columbus: Ohio State University Press, 148–163.
Andersson, Roger (toim.). 2014. *Heliga Birgittas texter på fornsvenska*. Bok 1 / Birgittas uppenbarelser. Stockholm: Sällskapet Runica et mediævalia.
Anttila, Raimo. 1989. *Historical and Comparative Linguistics*. Amsterdam: John Benjamins Publishing Company.
Auerbach, Eric. 2018. *Maailmankirjallisuuden filologia*. [*Gesammelte Aufsätze zur romanischen Philologie*, 1967.] Suom. Harry Lönnroth & Taina Vanharanta. Tampere: Vastapaino.
Carroll, Ruth, Peikola, Matti, Salmi, Hanna, Varila, Mari-Liisa, Skaffari, Janne & Hiltunen, Risto. 2013. Pragmatics on the page: Visual text in late medieval English books. *European Journal of English Studies* 17:1, 54–71.

Driscoll, Matthew James. 2010. The words on the page: Thoughts of philology, old and new. Julkaisussa: Quinn, Judy & Lethbridge, Emily (toim.) *Creating the Medieval Saga: Versions, Variability and Editorial Interpretations of Old Norse Saga Literature*. Odense: University Press of Southern Denmark, 87–104.

Gumbrecht, Hans Ulrich. 2003. *The Powers of Philology. The Dynamics of Textual Scholarship*. Urbana & Chicago: University of Illinois Press.

Gurd, Sean. 2010. Introduction. Julkaisussa: Gurd, Sean (toim.) *Philology and Its Histories*. Columbus: Ohio State University Press, 1–19.

Hogg, Richard. 1994. Linguistics, philology, chickens and eggs. Julkaisussa: Fernández, Francisco, Fuster, Miguel & Calvo, Juan Jose (toim.) *English Historical Linguistics 1992. Papers from the 7th International Conference on English Historical Linguistics*. Amsterdam: Johns Benjamins Publishing Company, 3–16.

Karlsson, Fred & Enkvist, Nils Erik. 2000. Kielitieteet. Julkaisussa: Tommila, Päiviö & Tiitta, Allan (toim.) *Suomen tieteen historia 2. Humanistiset ja yhteiskuntatieteet.* Porvoo: WSOY, 224–295.

Lönnroth, Harry. 2011. Kieli historiassa, historia kielessä – lingvistiikka, nordistiikka ja (uusi) filologia. *Tieteessä tapahtuu* 4–5, 26–31.

Lönnroth, Harry (toim.). 2017. *Philology Matters! Essays on the Art of Reading Slowly*. Leiden: Brill.

Marttila, Ville. 2014. *Creating Digital Editions for Corpus Linguistics: The case of Potage Dyvers, a family of six Middle English recipe collections*. Väitöskirja. Helsinki: University of Helsinki.

Melefors, Evert. 2013. *Senmedeltida kopparutvinning i Åtvidaberg speglad i räkenskaper från 1500-talets början*. Uppsala: Svenska fornskriftsällskapet.

Mäkilähde, Aleksi. 2019. *The Philological-Pragmatic Approach: A Study of Language Choice and Code-Switching in Early Modern English School Performances*. Väitöskirja. Turku: University of Turku. Saatavissa: http://urn.fi/URN:ISBN:978-951-29-7888-5.

Pollock, Sheldon. 2009. Future Philology? The Fate of a Soft Science in a Hard World. *Critical Inquiry* 35: 4, 931–961.

Pollock, Sheldon. 2015. Introduction. Julkaisussa: Pollock, Sheldon, Elman, Benjamin A. & Chang, Ku-ming Kevin (toim.) *World Philology*. Harvard: Harvard University Press, 1–24.

Sarasti-Wilenius, Raija. 2015. *Dear Brother, Gracious Maecenas: Latin Letters of the Gyldenstolpe Brothers (1661–1668)*. Helsinki: Academia Scientiarum Fennica.

Uusitalo, Harri. 2019. *Tausta, tekijä ja kieli: Filologinen tutkimus Aitolahden koodeksin lainsuomennoksesta*. Väitöskirja. Turku: Turun yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-29-7669-0.

Wegmann, Nikolaus. 2014. Philology – An Update. Julkaisussa: Bajohr, Hannes, Dorvel, Benjamin R., Hessling, Vincent & Weitz, Tabea (toim.) *The Future of Philology*. *Proceedings of the 11th Annual Columbia University German Graduate Student Conference*. Newcastle upon Tyne: Cambridge Scholars Publishing, 24–43.

# Luku 19 Nimistöntutkimus

*Paula Sjöblom*
https://orcid.org/0000-0002-7392-2300
*Terhi Ainiala*
https://orcid.org/0000-0003-4998-9983

**Mikä?**

Nimistöntutkimus eli onomastiikka tarkastelee erisnimiä monista eri näkökulmista. Nimiä voidaan tutkia osana kielijärjestelmää mutta myös omana systeeminään, joka toimii vuorovaikutuksessa ympäristön kanssa. Tästä syystä nimistöntutkimus kaipaa usein myös joltain muulta tieteenalalta, esimerkiksi historiasta, maantieteestä tai sosiaalipsykologiasta, saatavaa taustatietoa. Nimistöntutkimuksessa voidaan hyödyntää monia eri kielitieteen tutkimusmenetelmiä sen mukaan, minkä nimikategorian nimiä tutkitaan, mitä niistä halutaan selvittää ja millaiseen teoreettiseen taustaan tukeudutaan.

**Katso myös:**

Johdannon luku 1. Tutkimuksen käytäntö

Johdannon luku 4. Kielitieteen etiikka – hyvä tutkimus murroksessa

Luku 1 Kielitaju kielentutkijan työkaluna

Luku 2 Kysely

Luku 3 Haastattelu aineistonkeruun metodina sosiolingvistiikassa

Luku 5 Lingvistinen etnografia toimintayhteisöissä

Luku 6 Diskurssianalyysi

Luku 7 Laadullinen aineistopohjainen kielentutkimus

Luku 20 Tekstintutkimus

Luku 21 Sosiolingvistiikka

Luku 24 Psyko- ja neurolingvistiikka

## 1. Johdanto

Nimistöntutkimuksen eli onomastiikan tarkastelun kohteena ovat erisnimet eli proprit. Onomastiikka on siis eri asia kuin onomasiologia, joka tarkoittaa tietyn käsitteen tai käsitepiirin nimitysten tutkimusta (Tieteen termipankki s.v. onomasiologia). Onomasiologisen tutkimuksen kohteena ovat tavallisesti kielen yleisnimet eli appellatiivit, esimerkiksi eri maantieteellisiä käsitteitä tarkoittavien nimitysten kehittyminen (Laine 2007). Onomastiikkaakin voi toki harjoittaa onomasiologisesta näkökulmasta: on muun muassa tutkittu, millaisia nimimalleja 1800-luvulla syntyneelle kaupalliselle yritystoiminnalle kehittyi (Sjöblom 2006, 92).

Pohjimmiltaan onomastiikka on lingvististä tutkimusta, mutta tutkimuskohde ja tutkimuskysymykset vaativat usein moni- tai poikkitieteellistä lähestymistä. Onomastiikan juuret työntyvätkin tieteidenväliseen maastoon, sillä alun perin nimiä ryhdyttiin tarkemmin tutkimaan menneisyyden tutkimuksen arkeologian apuna. Vanhat paikannimet näet paljastavat yhtä ja toista ajoista, joista ei ole olemassa kirjallisia dokumentteja.

Nimistöntutkimus on vanhastaan jaoteltu henkilönnimitutkimukseen eli antroponomastiikkaan ja paikannimitutkimukseen eli

toponomastiikkaan. Viime vuosituhannen loppupuolelta tähän asti jaotteluun on ollut tapana lisätä kolmanneksi muiden nimien tutkimus, johon sisältyy hyvin erilaisia tutkimuskohteita eläinten nimistä yritysten nimiin tai kaunokirjallisuudessa esiintyvistä nimistä laivojen nimiin. Tällainen jaottelu perustuu nimikategorioihin, eli tällöin tutkimusta määrittää pääasiassa se, minkä lajin tarkoitteeseen nimi viittaa, eikä esimerkiksi tutkittavien nimien käyttökonteksti.

Tässä teoksessa onomastiikan metodeja ei esitellä nimikategorioittain, sillä yhteen ja samaan nimikategoriaan kuuluvia propreja on mahdollista tutkia monista eri näkökulmista ja monin eri menetelmin. Tarkoituksenamme on kuvata eri analyysitapoja, joista voi valita omaan tutkimuskohteeseensa ja tutkimusongelmaansa parhaiten soveltuvan metodin tai metodien yhdistelmän.

Luku 2 on yleiskatsaus proprien asemaan kielessä sekä nimistöntutkimuksen taustateorioihin. Luvussa 3 esittelemme eri luvuissa yleisimmät onomastiikassa käytössä olevat analyysimallit, ja luvun 4 yhteenvedossa kokoamme päälinjat ja luomme silmäyksen onomastiikan tulevaisuuden mahdollisuuksiin.

**Mihin?**

Onomastiikan tutkimuskohteena ovat kielen sanat, joilla viitataan vain yhteen tarkoitteeseen. Nimen perustehtävänä on identifioida entiteetti muista samankaltaisista. Nimien tarkoitteita voivat olla todelliset tai kuvitteelliset henkilöt, paikat, eläimet, ihmisen rakentamat tai tuottamat esineet ja asiat. Nimistöntutkimuksen avulla voidaan saada tietoa niin asutushistoriasta, identiteettien rakentumisesta, ihmisen ajattelusta, arvostuksista ja sosiaalisista suhteista kuin myös kielisysteemistä.

# 2. Tutkimuksen lähtökohtia

Sanaluokkana erisnimet kuuluvat suomen kielessä substantiiveihin. Ne taipuvat luvussa ja sijassa kuten yleisnimetkin ja voivat toimia substantiivilausekkeissa edussanoina. Proprien ja appellatiivien välinen raja on sumea, mutta olennainen ero näiden kahden substantiiviryhmän kesken on se, että appellatiivit ilmaisevat luokan, johon samankaltaiset entiteetit kuuluvat (esim. *kaupunki*), mutta proprit nostavat esiin eli identifioivat ja erottavat yksilön muiden samaan luokkaan kuuluvien entiteettien joukosta (esim. *Porvoo*) (ISK § 553).

Normitetussa kirjakielessä proprit kirjoitetaan isolla alkukirjaimella, mutta kirjoitetuista teksteistä nimiaineistoa keräävän pitää muistaa, että alkukirjain osoittaa vain kirjoittajan tulkinnan eikä rajaa sanan kuulumista propreihin tai appellatiiveihin. Proprin tunnistamiseen tarvitaan siis muita kriteerejä kuin sanan alkukirjaimen koko.

Erisnimellä on aina jokin muoto, äänneasu, joka viittaa johonkin kielenulkoiseen entiteettiin, joko todelliseen tai kuvitteelliseen. Tätä entiteettiä kutsutaan nimen *referentiksi*. Erisnimen tunnistamisen ensimmäinen kriteeri on, että nimi viittaa vain yhteen referenttiin; nimi on siis monoreferentiaalinen eli yksiviitteinen sana.

Käytännössä asia ei kuitenkaan ole näin yksinkertainen. Useimmat tuntemamme henkilönnimet viittaavat moniin eri henkilöihin. Etu- ja sukunimen yhdistelmäkään ei ollenkaan aina ole ainutlaatuinen. Saman voi todeta monista paikannimistä: esimerkiksi *Mustalammet* ja *Saarijärvet* toistuvat kartalla tiheään tässä tuhansien järvien maassa (ks. Nimisampo). Entä sitten sellaiset nimet kuin *Adidas* tai *Ikea*? Eivätkö ne olekaan nimiä, kun *Adidas* näkyy kaupan hyllyllä monen tuotteen kyljessä tai kun *Ikea* lukee rakennuksen seinässä niin Raisiossa kuin Tukholmassakin?

Kaikki edellä mainitut ovat kuitenkin propreja. Ratkaisu ongelmaan on konteksti eli esiintymisympäristö. Erisnimi on kontekstissaan yksiviitteinen. Vuorovaikutuksen osapuolilla on yleensä yhteinen käsitys siitä, kehen tai mihin nimellä viitataan. Tämä käsitys voi perustua lauseyhteyteen (esim. *Pääsikö* ***teidän Jenni*** *yliopistoon?*) tai asiayhteyteen (esim. *Naapurini tytär sai opiskelupaikan yliopistosta.* ***Jenni*** *oli sinnikäs; vasta*

*kolmas kerta tuotti tuloksen*). Kaupalliset tuotteiden ja yritysten nimet taas ovat nimiä, kun ne viittaavat yhteen ainutlaatuiseen, muista tuotteista tai yrityksistä eroavaan ideaan – konkreettiset esineet tai rakennukset ovat vain tämän idean toisintoja.

Toisaalta kontekstin pohjalta sanan voi tulkita nimeksi, vaikka sitä yleensä pidettäisiin luokittelevana appellatiivina. Jotta kuulija ymmärtäisi oikein lauseen *Nähdään kirkolla*, täytyy hänen tunnistaa sana *kirkko* yhteen tiettyyn kirkkoon viittaavaksi. Vaikka sellaiset sanat kuin *kirkko*, *risteys*, *amis* tai *äiti* ovat joissain konteksteissa tulkittavissa yksiviitteisiksi propreiksi, ei niitä onomastiikassakaan yleensä pidetä niminä.

Sana on mahdollista tulkita kontekstissaan yksiviitteiseksi ja siten propriksi vaikkapa taivutuksensa vuoksi. Lihavoitu *yliopisto*-sana seuraavassa esimerkissä viittaa yksiselitteisesti yhteen tiettyyn yliopistoon, tässä kontekstissa Turun yliopistoon:

”*Turun maakuntamuseon lisäksi Ristimäen aineistoa käydään läpi* ***yliopistolla*** *sekä muissa Suomen yliopistoissa ja tutkimuslaitoksissa*.” (Turun Sanomat 9.10.2017).

Yksi erisnimen tunnistamisen kriteeri on sen muoto. Monet proprit ovat äänneasultaan läpinäkymättömiä, opaakkeja, mikä tarkoittaa, ettei ilmaus sisällä kielenpuhujille tuttuja kielen sanoja. Tällaisia ovat esimerkiksi useimmat etunimet (*Maija*, *Eero*), lukuisat vanhat paikannimet (*Naantali*, *Päijänne*, *Viro*) ja monet yritysten tai tuotteiden nimet (*Linux*, *Abloy*). On varsin todennäköistä, että tekstissä esiintyvä muodoltaan tuntematon ilmaus on erisnimi, ellei se ole esimerkiksi uusi lainasana. Toisaalta myös monet läpinäkyvät eli transparentit ilmaukset voidaan tunnistaa nimiksi, koska niillä tuskin on kielessä täsmälleen samanlaista appellatiivista homonyymia (esim. *Rautavaara*, *Housulampi*) tai ne poikkeavat appellatiiveista prosodisesti, morfologisesti tai syntaktisesti (*Mustasaari* : *Mustasaaressa*; *Varpu* : *Varpun*). (Kiviniemi 1975.)

Filosofit ovat olleet kiinnostuneita nimistä ja nimien olemuksesta jo antiikin ajoista lähtien. Muun muassa kysymystä proprin ja

appellatiivin välisestä rajasta on kielifilosofiassa pohdittu yhä uudelleen ilman lopullista ratkaisua (esim. Frege 1892; Mill 1906; Jespersen 1958 [1924]; Gardiner 1940; Russell 1956; Kripke 1972). Varsinainen onomastiikka syntyi kuitenkin 1800-luvulla lähinnä arkeologian, historian ja kielihistorian tarpeisiin, minkä vuoksi tutkimus kohdistui ensin lähes yksinomaan paikannimiin. Nimistä ei oltu kiinnostuneita niinkään kielen sanoina vaan muistomerkkeinä menneiden aikojen ihmisistä, heidän tavoistaan ja kielestään. Ensimmäinen suomalainen nimistöntutkimuksen väitöskirja käsitteli henkilönnimiä: A. V. Forsmanin (1891) tutkimuskohteena olivat muinaissuomalaiset henkilönnimet. Etymologisen henkilön- ja paikannimitutkimuksen avulla onkin pystytty selvittämään merkittävällä tavalla menneisyyteen liittyviä kysymyksiä.

Suomalaisen etymologisen paikannimitutkimuksen uranuurtaja oli Helsingin yliopiston nimistöntutkimuksen dosentti ja Suomen nimiarkiston johtaja Viljo Nissilä (1903–1986). Hän keräsi ja tutki kymmeniätuhansia nimiä eri maantieteellisiltä alueilta Karjalan kannakselta Satakuntaan 1930-luvulta alkaen (ks. esim. Nissilä 1939; 1962; 1975). Nissilän ansiosta suomalainen onomastiikka saavutti nopeasti kansainvälisen tason. (Virtaranta 1973.)

Lingvistisenä tutkimusalana onomastiikka on 1900-luvulta asti saanut vaikutteita kulloinkin vallalla olleista kieliteoreettisista näkemyksistä. Kielen rakenteen abstrahoiminen toisistaan erillisiksi tasoiksi ja semantiikan ja pragmatiikan irrottaminen rakenteen tarkastelusta johti onomastiikassa siihen, että nimistöä – jota oli semanttisesti hankala verrata muuhun sanastoon – keskitytttiin tutkimaan omana itsenäisenä systeeminään ja omista typologisista lähtökohdistaan. Tästä näkökulmasta nimiä on tietenkin edelleen mahdollista ja tarpeellista tutkia.

1900-luvun jälkipuolen sosiolingvistiset, systeemis-funktionaaliset, psykolingvistiset ja kognitiiviset kieliteoriat ovat havaittavissa, vaikka eivät läheskään aina eksplikoituina, 1960-luvun jälkeisessä onomastiikassa. Yhä enenevässä määrin nimiä tarkastellaan osana koko kielisysteemiä, tutkitaan nimeämistä ja nimien käyttöä niin sosiaalisena kuin myös kognitiivisena ilmiönä ja pohditaan nimien asemaa osana viestintää ja vuorovaikutusta. Onomastiikan alalta on 1970-luvulta lähtien

valmistunut kiihtyvään tahtiin väitöskirjoja, joiden tutkimuskohteet ja -näkökulmat ovat haaroneet moniin eri suuntiin.

Nimet ovat siis osa kieltä, mutta ne ovat myös kieltenvälisiä, koska niitä käytetään usein sellaisenaan kielestä riippumatta: nimet lainautuvat kielestä toiseen helposti. Nimien semanttinen erityisluonne aiheuttaa sen, että niiden tutkimuksessa on syytä ottaa huomioon myös kieltä laajempi konteksti. Nimet ovat yhteiskunnan toiminnan ja ihmisten itsensä kannalta erittäin tärkeitä kielen sanoja. Tutkimuskohteensa ja -näkökulmansa mukaan tutkijan on välttämätöntä perehtyä esimerkiksi tutkimansa alueen maantieteeseen, tutkimiaan nimiä koskevaan lainsäädäntöön, talouselämän realiteetteihin, sosiaalisiin näkökulmiin tai identiteettikysymyksiin.

# 3. Nimistöaineistot ja niiden analyysi

Tässä luvussa esitellään kymmenen erilaista nimistöntutkimuksen analyysimallia. Esiin nousevat typologinen analyysi, kognitiivinen merkitysanalyysi, leksikaalis-etymologinen analyysi, kontaktionomastiikka, sosiolingvistinen variaationanalyysi, nimet vuorovaikutuksessa, kansanonomastiikka, tekstianalyyttinen ja diskurssianalyyttinen metodi, nimimaisematutkimus ja multimodaalinen analyysi sekä suuret korpukset ja niiden määrällinen analyysi. On syytä muistaa, että nimistöä voi käytännössä tutkia monilla niillä metodeilla, joita tässä teoksessa esitellään. Nimistön tietynlainen erikoisluonne lingvistisenä kohteena on kuitenkin otettava huomioon metodia sovellettaessa, ja tietyt metodit ovat toistaiseksi vakiintuneempia onomastiikassa kuin toiset.

## 3.1. Typologinen analyysi

Nimiä ja nimistöjä tutkittaessa typologinen analyysi jää vain harvoin syrjään. Vaikka ensisijaisena tutkimuskohteena olisi jokin muu, on tutkijan

kuitenkin ensin otettava selvää nimien typologiasta eli siitä, millä tavoin ja millä perustein nimet jaetaan eri nimenosiin. Tässä luvussa esitellään keskeisimmät nimistön rakenneanalyysit: syntaktis-semanttinen, funktionaalis-semanttinen ja henkilönnimien rakenneanalyysi.

**Syntaktis-semanttinen rakenneanalyysi**. Syntaktis-semanttinen rakenneanalyysi on perinnäisten paikannimien rakenteen ja semantiikan ymmärtämiseksi luotu malli (Kiviniemi 1975; Zilliacus 1966). Metodin kehittämisen taustalla vaikutti transformatiivis-generatiivinen kielioppi (ks. Kiviniemi 1987). Mallissa pyritään nimenantajan näkökulmasta selvittämään, mitkä asiaperusteet ovat löydettävissä nimeämisen taustalta. Nimet jaetaan ensin syntaktisiin rakenneosiin eli nimenosiin. Ne taas ovat löydettävissä nimien ns. syvärakenteiden avulla, jotka ovat nimien pintarakenteiden taustalla olevia lauseparafraaseja. Esimerkiksi lammen nimen *Mustalampi* lauseparafraasi voi olla 'lampi, jonka vesi on mustaa'. Näin lauseparafraasi paljastaa nimen kaksi syntaktista rakenneosaa (*musta* + *lampi*) ja niiden syntaktisen suhteen.

Syntaktinen nimenosa ilmaisee yhden tarkoitetta luonnehtivan seikan eli yhden paikalle ominaisen piirteen. Näitä nimenosia voi olla kolmenlaisia. Ensiksikin nimenosa voi olla paikan lajia ilmaiseva (*Musta/**lampi***). Kaksiosaisen yhdysnimen jälkiosa eli nimen perusosa on yleensä juuri esimerkin kaltainen paikkaa luonnehtiva appellatiivi. Toiseksi nimenosan funktio voi olla erityispiirrettä ilmaiseva (***Musta**/lampi*), jolloin nimi ilmaisee jotakin nimen tarkoittamasta paikasta. Useimmiten erityispiirrettä ilmaisevat nimenosat esiintyvät nimissä määriteosina eli ovat yhdysnimen alkuosina. Kolmanneksi nimenosan funktio voi olla nimittävä, ja tällainen funktio voi olla joko määrite- tai perusosalla. Nimittäviä ovat sellaiset propriset eli nimen sisältävät nimenosat, joilla ei ole erityispiirrettä ilmaisevaa funktiota. Nimenosa ei siis ilmaise paikan lajia eikä kuvaile paikkaa jonkin sen erityispiirteen osalta. Tällaisia ovat esimerkiksi lainanimet (*Porvoo* vrt. *Borgå*), joita lainanottajien kannalta voidaan kuvailla sisällöttömiksi.

Rakenneanalyysin seuraava vaihe on erityispiirrettä ilmaisevien nimenosien luokittelu sekä lausesemanttisesti että sanasemanttisesti. Lausesemanttisessa luokittelussa selvitetään nimien nimeämisperusteet. Ne ovat kielenulkoiseen todellisuuteen liittyviä seikkoja, kuten

paikan sijainti tai ominaisuus. Sanasemanttinen luokittelu puolestaan ilmaisee sen leksikaalisen kategorian, johon kuuluvalla ilmauksella nimeämisperuste on ilmaistu. Esimerkiksi nimessä *Mustalampi* määriteosa *Musta* kertoo paikan ominaisuudesta ja leksikaalisena kategoriana se on adjektiivi. Mallissa lause- ja sanasemanttisella luokittelulla on omat kirjainkoodit. (Kiviniemi 1990.)

Mallia on käytetty laajalti perinnäisen paikannimistön tulkinnassa. Tällöin aineistona on voinut olla esimerkiksi jonkin kylän koko nimistö, ja analyysin tuloksena on yleensä pystytty tulkitsemaan eli sijoittamaan luokitteluun suuri osa paikannimistä, jopa yli 90 prosenttia. On kuitenkin otettava huomioon, että kaikkien nimien nimeämisperusteista ei voi saada varmaa vastausta. Lisäksi luokittelu ei sovi muun kuin perinnäisen paikannimistön analyysiin. Kaavanimistöä ei siis pidä lähteä analysoimaan tätä mallia käyttäen.

Nimenosan syntaktis-semanttinen funktio voi olla:

A = paikan lajin ilmaiseva (*Musta/**lampi***)

B = nimittävä (*Porvoo*)

C = erityispiirrettä ilmaiseva (***Musta**/lampi*)

Koodin alussa oleva numero ilmaisee nimenosien lukumäärän.

Esimerkkejä:

*Musta/lampi* ’lampi, jonka vesi on mustaa’ 2 CA

*Mustalampi* ’(talo) Mustalammen rannalla’ 1 C

*Iso / Mustalampi* ’isompi *Mustalampi*-nimistä lammista’ 2 CB

*Katinhäntä* ’Katinhännän muotoinen (pelto)’ 1 C

*Niemi* ’niemi’ 1 A

**Funktionaalis-semanttinen rakenneanalyysi**. Perinnäisten maaseudun paikannimien tutkimiseen kehitetty syntaktis-semanttinen malli ei sovellu kaikkien nimikategorioiden typologiseen analyysiin. Perinnäiset paikannimet on yksinkertaista jakaa syntaktiselta pohjalta perus- ja määriteosiin, ja nimeämisen perusteet on periaatteessa mahdollista jäljittää nimen syntyhetkellä vallinneisiin olosuhteisiin. Sitä vastoin monet muut nimikategoriat ovat rakenteeltaan paljon kompleksisempia (esim. asunto-osakeyhtiön nimi *Turun työläisten as.oy Itärinne*) tai ne perustuvat aivan toisenlaisiin seikkoihin kuin tarkoitteen sijaintiin, ominaisuuteen tai käyttöön (esim. kaupunkinimistössä tavalliset muistonimet, kuten *Sibeliuksenkatu*).

Yritysten nimien rakenneanalyysiin kehitetty funktionaalis-semanttinen analyysimalli (Sjöblom 2006) soveltuu metodiksi erityisesti, kun halutaan tarkastella nimiä, jotka ovat syntyneet tietoisen ja harkitun nimeämisprosessin kautta ja joilla on useita eri käyttöfunktioita. Tällaisten nimien rakenne voi olla erittäin monimutkainen: nimet ovat usein monisanaisia, eikä niistä ole mahdollista tunnistaa elementtejä, jotka olisivat selkeässä perus- ja määriteosasuhteessa toisiinsa. Monilla nimillä on virallisen, johonkin rekisteriin kirjatun muodon lisäksi jokin lyhyempi käyttömuoto.

Funktionaalis-semanttisessa analyysissa selvitetään aluksi, millaisista funktionaalisista osista kompleksiset nimet koostuvat. Mallissa nimenosat määritellään nimikategorian omista lähtökohdista käsin. Yritysten nimet, tuotteiden nimet tai verkkopalvelujen käyttäjänimet ovat kaikki rakenteeltaan erilaisia. Nimitypologian selvittämiseksi onkin perehdyttävä suureen joukkoon saman kategorian nimiä, joita vertaamalla funktionaaliset osat paljastuvat. Koska nimen perustehtävä on aina yksilöinti, on kaikissa nimissä välttämättä jokin osa, jota voi pitää nimen varsinaisena yksilöivänä osana.

Esimerkiksi viralliset, rekisteröidyt yritysnimet voivat sisältää neljänlaisia nimenosia: yritysmuodon ilmaisevan osan, liikeidean ilmaisevan osan, yksilöivän osan ja täydentävää informaatiota antavan osan. Näiden

nimenosien järjestys voi vaihdella, eikä niiden keskinäistä syntaktista suhdetta voi määritellä helposti. Kaikissa yritysnimissä ei myöskään ole kaikkia neljää osaa. Vain yksilöivä osa on pakollinen – myös lain mukaan.

Esimerkki rekisteröidyn yritysnimen nimenosista:

| *Kauneushoitola* | *Päivänpaiste* | *tmi* | *Tarja Tappurainen* |
|---|---|---|---|
| liikeideaosa | yksilöivä osa | yritysmuoto-osa | täydentävä osa |

Analyysin toisessa vaiheessa selvitetään, minkätyyppisistä ilmauksista eri nimenosat, etenkin yksilöivät nimenosat, rakentuvat eli millaisia sana-aineksia ne sisältävät. Ne voivat sisältää esimerkiksi henkilön- tai paikannimiä tai muita erisnimiä (*Wärtsilä, Otava*), appellatiiveja tai muiden sanaluokkien sanoja (*Kone, Metso, Valio*), useammasta sanasta koostuvia lausekkeita, fraaseja, tekosanoja, kirjainlyhenteitä tai numeroita (*Teollisuuden Voima, Scanfil, UPM, St1*). Lisäksi monien nimikategorioiden kannalta myös nimien ja nimenosien kielen määrittely sekä nimien ortografia voi olla merkittävää.

Typologisen analyysin tavoitteena on saada aineistosta yleiskuva, jonka perusteella voidaan selvittää vastausta kysymykseen, miksi nimet ovat sellaisia kuin ovat. Funktionaalis-semanttista analyysia voikin jatkaa tutkimalla tarkemmin nimien semantiikkaa ja funktioita (ks. luku 3.2.).

Funktionaalis-semanttista mallia on sovellettu etenkin kaupallisten nimien (yritys- ja tuotenimet) ja verkkoympäristön käyttäjänimien (esim. Hämäläinen 2016) typologiseen analyysiin. Koska malli on joustava ja siinä otetaan huomioon nimikategorian ominaispiirteet, sitä voi kokeilla myös muiden sellaisten nimien luokitteluun, joita on vaikea sovittaa syntaktis-semanttisen rakenneanalyysin kaavaan. Esimerkiksi ravihevosten nimien[1], koirien kennelnimien tai vaikkapa lastenkirjallisuuden nimien[2] tutkimuksessa mallia voi mahdollisesti hyödyntää. Kaikkien nimien typologiaa ei kuitenkaan voi selvittää tälläkään analyysimenetelmällä. Esimerkiksi erilaisten taideteosten (kirjojen, taulujen, elokuvien jne.) nimiä ei välttämättä pysty jaottelemaan funktionaalisiin saati syntaktisiin osiin, vaan nimet saattavat muodostaa yhden

erottamattoman kokonaisuuden. Tällaisillakin nimillä on omat typologiset piirteensä, mutta niitä on tutkittu toistaiseksi erittäin vähän.

**Henkilönnimien rakenneanalyysi**. Henkilönnimet voidaan jakaa karkeasti kahteen pääryhmään, yksilöihin viittaaviin nimiin (esim. etunimet) ja kollektiivisiin nimiin (esim. sukunimet). Eri kulttuureissa henkilönnimisysteemit ovat hyvin erilaisia jopa Euroopan mittakaavassa; esimerkkinä vaikkapa Islannin etunimi + isännimi -systeemi ja Espanjan etunimi + isän sukunimi + äidin sukunimi -systeemi. (Lisätietoa henkilönnimisysteemeistä, ks. Ainiala, Saarelma & Sjöblom 2008.) Henkilönnimien typologian tutkimiseen on olemassa eri menetelmiä.

Etunimiä on perinteisesti luokiteltu etymologisin perustein eri-ikäisiin kerrostumiin. Muinaissuomalaisista nimistä tiedetään melko vähän, koska kirjalliset lähteet ovat harvassa ja kristinuskon myötä nimistö vaihtui meillä pikkuhiljaa kokonaan kristilliseksi saamelaisasutusta lukuun ottamatta. Tärkeitä tutkimuslähteitä ovat meillä kuitenkin olemassa olevat sukunimet ja paikannimet, joissa on säilynyt paljon vanhaa henkilönnimistöä.

Kristillisperäisiksi nimiksi luetaan paitsi raamatulliset (Uuden ja Vanhan testamentin) nimet myös muiden eurooppalaisten kansojen alkuperäiset pakanalliset nimet, joista tuli pyhimysten nimiä (esim. *Erik*, *Kerttu* < *Gertrud*). 1800-luvulta lähtien etunimistöön on tullut jälleen vanhoja suomalaisnimiä (*Väinö, Kyllikki*) ja uusia omakielisiä nimiä (*Urho, Rauha*). Uusimpiin nimikerrostumiin kuuluu suomenkielisiä nimiä, muista kielistä lainattuja nimiä sekä uniikkeja sepitteitä.

> Katolisen kirkon keskiaikainen pyhimyskalenteri sisälsi nykyisin tuntemaamme nimipäiväkalenteriin verrattuna paljon vähemmän eri nimiä. Lisäksi useimmat pyhimysnimet olivat miesten nimiä. Lasten nimeämisessä päädyttiin usein samaan nimeen. Keskiajalla *Johannes* oli Euroopassa jossain vaiheessa niin tavallinen nimi, ettei se erotellut enää riittävästi edes pienen yhteisön yksilöitä toisistaan. Sen vuoksi nimestä otettiin käyttöön useita eri variantteja. Suomessa *Johanneksen* variantteja ovat mm. *Jussi*, *Janne*, *Juha*, *Juho*, *Jani*, *Juhani*, *Hannes* ja *Hannu*.

Viimeaikaisessa etunimitutkimuksessa käyttökelpoiseksi on osoittautunut *muotin* käsite (Sartjärvi 2013; vrt. ISK § 148), jonka rinnalla voitaisiin käyttää myös *konstruktion* käsitettä. Etunimet muodostavat samaloppuisia rypäitä, jotka toimivat mallina uusille etunimille. Esimerkiksi nimet *Eija, Seija, Maija, Aija* ja *Tuija* mukautuvat etunimimuottiin *(C)V+ija*, jota voidaan käyttää edelleen uusien nimien muodostamiseen (esim. *Beija, Freija, Paija, Roija, Toija, Viija*). Naistennimissä etunimimuottien vaikutus näkyy selvemmin kuin miestennimissä, mikä johtuu siitä, että tytöille annetaan ylipäätään enemmän erilaisia nimiä kuin pojille. Poikia nimetään useammin suvun nimien mukaan, minkä vuoksi pojilla käytössä oleva nimivarasto ei ole yhtä suuri kuin tytöillä.

Suomalaisia sukunimiä voidaan luokitella rakenteensa perusteella kantasana- ja yhdyssanatyyppisiin sekä johtimellisiin nimiin. Kantasanatyyppiset sisältävät suomen kantasanoja, kuten *Laine*, tai muita yksiosaisia ilmauksia, kuten etunimiä (*Pentti*), niistä johdettuja hypokorismeja (*Rusi* < *Erasmus*) tai uusmuodosteita, kuten *Vento*. Yhdyssanatyyppiset sukunimet ovat yleensä joko vanhoja talonnimiä (*Saarikoski*) tai uusia, sukunimiksi sepitettyjä muodosteita (*Vuorenjuuri*). Yleisimmät suomalaiset sukunimijohtimet ovat -*lA*, -*nen*, -*mO*, -*iO*, -*mA* ja -*stO*. (Suomalaisesta sukunimisysteemistä lisää ks. Paikkala 2004.)

Epävirallista henkilönnimistöä ovat lisä- ja pilkkanimet (*Pontikka-Topi, Hönö*), erilaiset hellittelynimet (*Annuliini*) ja hypokorismit eli virallisesta nimestä johdetut nimiasut (*Pirkko* > *Pike*). Analyysissa on tapana jakaa nämä nimet sekundaareihin ja primaareihin lisänimiin. Sekundaarien lisänimien pohjana ovat ihmisten viralliset etu- ja sukunimet (*Kalervo* > *Kale, Peltonen* > *Peltsi*), ja primaarit lisänimet perustuvat johonkin muuhun ihmiseen liittyvään seikkaan, kuten ulkonäköön, ammattiin, käyttäytymiseen tai tapahtumaan. Sekundaarien lisänimien muodostamiseen käytetään yleisiä sananmuodostuskeinoja, kuten typistämistä, johtamista ja äänteellistä muuntelua, ja apua on erityisesti slangijohtimien ja -johtamissääntöjen tuntemuksesta. (Ainiala, Saarelma & Sjöblom 2008.)

Henkilönnimien typologian tutkimukseen aineistoa voidaan kerätä esimerkiksi eri aikojen nimipäiväkalentereista (ks. Saarelma 2006) sekä Väestörekisterikeskuksen etu- ja sukunimirekistereistä.

## 3.2. Kognitiivinen merkitysanalyysi

Perinnäisissä paikannimissä nimen syntyhetken nimeämisperuste ja nimen silloinen sisältö eivät myöhemmän käytön kannalta ole merkityksellisiä, vaan paikannimi yksilöi tietyn paikan riippumatta siitä, tunnistavatko kielenpuhujat nimeämisperusteen vai eivät. Monissa muissa nimikategorioissa nimeen sisältyvien sanojen merkitys voi kuitenkin olla nimen käyttäjien kannalta oleellinen. Esimerkiksi monien yritysten ja tuotteiden nimet tai vaikkapa henkilöiden epäviralliset nimet toimivat parhaiten tehtävässään, kun niiden sisältö on tunnistettavissa ja tavalla tai toisella tulkittavissa. Alkuperäinen nimeämisperuste ei ole tällaisten nimien käytön kannalta oleellinen, vaan oleellisempaa on tiedostaa nimen välittämät viestit. Tällaisiin tapauksiin tarvitaan metodi, joka ottaa huomioon identifioinnin lisäksi myös nimen muut mahdolliset tehtävät. Näihin tehtäviin nimen sisältö voi vaikuttaa ratkaisevasti.

Kognitiivinen merkitysanalyysi on malli, joka on syntynyt funktionaalis-semanttisen rakenneanalyysin (ks. luku 3.1.) täydennykseksi (Sjöblom 2006). Malli perustuu kognitiiviseen kielinäkemykseen ja tarkastelee nimiä osana koko kielisysteemiä sekä suhteessa ihmisen tietorakenteisiin ja kommunikaatiotilanteisiin. Nimiä voi pitää muiden kielten sanojen tapaan symbolisina yksiköinä, joilla on muoto ja merkitys. Merkitys tarkoittaa proprin muodon aktivoimia linkkejä eri kognitiivisiin alueisiin eli merkitysyhteyksiä, joista osa on subjektiivisia mutta osa kielenpuhujille yhteisiä.[3]

Kognitiivisen merkitysanalyysin tavoitteena on selvittää funktionaalis-semanttisen analyysin perusteella löydettyjen nimenosien semanttiset piirteet. Erityisen kiinnostavaa on tutkia yksilöivän osan sisältämän kielellisen ilmauksen merkitystä ja suhteuttaa se nimen tarkoitteeseen. Toisin sanoen menetelmällä kuvataan, millaisessa merkityssuhteessa nimi on tarkoitteeseensa nähden. Tämä vaatii tietysti nimen tarkoitteesta riittävästi taustatietoa.

Useimmiten nimenosan merkityssuhde nimen tarkoitteeseen on joko suora tai epäsuora. Suora merkityssuhde tarkoittaa sitä, että nimenosa kertoo tarkoitteesta asioita suoraan: esimerkiksi nimessä *Kukkakauppa Virtanen* liikeideaa ilmaiseva nimenosa *Kukkakauppa* ilmaisee yrityksen

toimialan ja *Virtanen* omistajan tai perustajan sukunimen. Epäsuora merkityssuhde taas tarkoittaa sitä, että nimenosan sisältämät ilmaukset ovat yhteydessä nimen tarkoitteeseen jonkin assosiaation kautta eli joko metonyymisesti kosketusassosiaation kautta (esim. kampaamon nimi *Rajut rastat*) tai metaforisesti yhtäläisyysassosiaation kautta (esim. päiväkodin nimi *Satumetsä*). Joskus nimenosan suhde tarkoitteeseen voi olla myös kompleksisempi tai nimenosan muodon ja tarkoitteen välillä ei ole lainkaan välitöntä merkityssuhdetta, vaan nimeen kiinnittyvät merkitykset rakentuvat välillisesti ajan kuluessa. (Sjöblom 2006.)

Kognitiiviseen analyysimalliin kuuluu myös nimien sosiokulttuuristen funktioiden tarkastelu. Tällöin pohditaan vastausta siihen, miksi nimet ovat sellaisia kuin ovat. Taustalla on ajatus, että merkitysten tuottaminen ei ole koskaan itseisarvo vaan merkityksellä on aina jokin funktio. Esimerkiksi yrityksen nimi saattaa integroida yrityksen muiden saman alan yritysten joukkoon tai jonkin kaupunginosan epävirallinen, lyhyt nimi voi helpottaa nimen käyttöä. Nimen tehtävänä voi olla myös nimen tarkoitteen, kuten jonkun henkilön tai jonkin paikan, pilkkaaminen. Tavallisimpia nimien funktioita ovat personoiva, houkutteleva, käytännöllinen, informatiivinen, integroiva, esteettinen ja pilkkaava funktio. Nimikategoriasta riippuen muitakin funktioita voi löytää. Monet nimet toimivat samanaikaisesti useassa eri funktiossa. Kognitiivista mallia on sovellettu muun muassa laivojen (Schybergson 2009) ja ajoneuvojen epävirallisten nimien (Seppälä 2013) tutkimiseen.

Erotuksena nimien typologiseen analyysiin kognitiivinen näkökulma mahdollistaa nimien merkityksen synkronisen tarkastelun riippumatta diakronisesta nimeämisen motiivista. Kognitiivisessa analyysissa näkökulma on kieliyhteisössä, nimenkäyttäjissä ja heidän tulkinnoissaan, kun taas varsinkin syntaktis-semanttisessa analyysissa näkökulma on nimeämishetkessä, nimenantajassa ja nimeämisperusteissa. Kaikissa edellä kuvatuissa metodeissa lähestytään kuitenkin nimistöä systeeminä, joka tarjoaa kieliyhteisölle ja nimenantajille mallin uusien nimien muodostamiseen. (Sjöblom 2004; 2006, 62–74.)

## 3.3. Leksikaalis-etymologinen analyysi

Leksikaalis-etymologisessa analyysissa selvitetään nimen alkuperää eli sitä, mistä ja mitä merkitsevistä kielellisistä elementeistä nimi koostuu ja millainen on ollut nimen alkuperäinen asu. Aivan ensimmäiseksi tulee tutkia, mitä kieltä nimi alkuaan on ollut. Monien taustaltaan hämärien nimien kohdalla kysymys voi jäädä lopullista vastausta vaille (*Imatra, Päijänne, Hyrkäs, Kuronen*), mutta perusteltuja vaihtoehtoja on silti yleensä löydettävissä. Lisäksi nimi pyritään ikäämään niin tarkoin kuin mahdollista. Iän määrittämisessä apuna ovat varhaiset asiakirja- ja karttamerkinnät, jotka kertovat ainakin sen, milloin nimi viimeistään on ollut käytössä.

Jos nimi koostuu tutuista kielen sanoista eli on läpinäkyvä, on nimen alkuperän selvittäminen huomattavasti helpompaa kuin läpinäkymättömien nimien kohdalla. Läpinäkyvänkin nimen kohdalla on kuitenkin perusteltua lisäksi selvittää, mikä nimen nimeämisperuste on ollut (ks. syntaktis-semanttinen rakenneanalyysi luvussa 3.1.) ja mitä nimen lekseemit tai muut elementit juuri tässä nimessä merkitsevät. Esimerkiksi *Kirkkokallio*-nimen nimeämisperusteena voi olla esimerkiksi 'kallio, jolla sijaitsee kirkko' tai 'kallio, joka muistuttaa kirkkoa'. *Akanniemi*-nimi puolestaan on edellistä monitulkintaisempi, koska emme tiedä, viittaako *akka* nimessä esimerkiksi vanhaan naiseen, isoäitiin tai voiko se olla mytologinen, jolloin viitattaisiin paikan rituaaliseen luonteeseen (ks. Mallat 2007).

Nimeen sisältyvät sanat ja muut kielelliset ainekset voivat olla sellaisia, ettemme enää tunnista nimen sisältöä. Tällaiset läpinäkymättömät nimet voivat sisältää aineksia, jotka ovat jääneet pois paikallisesta kielestä. Esimerkiksi kainuulaisen kylännimen *Neittävä* taustalla lienevät murteissa vanhastaan tunnetut verbit *neitää* ja *neityä* 'vuotaa, tihkua läpi; kostua'. Alue on kahden vesistön välisellä alavalla maalla, joten merkitykseltään nimi sopii paikalle hyvin. Tutkijan tulee siis hankkia laaja-alaisesti tietoa kielihistoriasta ja murteista sekä lisäksi maantieteen, paikallishistorian ja muiden lähitieteiden avulla selvittää, millaisia nimetyt kohteet nimenantoaikoina ovat voineet olla.

Nimi voi olla läpinäkymätön eli hämärä siksikin, että siinä on aikojen kuluessa tapahtunut erilaisia rakenteellisia ja äänteellisiä muutoksia. Paikannimen määriteosassa on esimerkiksi voinut tapahtua

loppuheitto (*Hailuoto* < *Haililuoto*; *haili* 'silakka', *Koijärvi* < *Koivujärvi*), ja nimen perusosa taas on voinut ikään kuin kulua (*Mustajoki* > *Mustio*). Nimen alkuperäisen tai ainakin vanhemman asun etsinnässä vanhat asiakirja- ja karttamerkinnät voivat olla arvokkaita. Esimerkiksi hämäläinen *Mainiemi*-nimi on mainittu 1500-luvun alun asiakirjassa asussa *Mayenemi*, joka todennäköisesti on monien alueen muiden *maja*-nimien lailla lyhentynyt *majava*-nimestä ja viittaa majavien pesimiseen alueella. Nimien alkuperää tutkittaessa pitääkin tuntea myös alueen ja yhteisön muu nimistö ja sen rakenne, jotta voi päätellä, millaiset leksikaaliset, rakenteelliset ynnä muut kielelliset valinnat sekä nimeämisperusteet ylipäänsä ovat mahdollisia.

Nimistöntutkijan on lisäksi otettava huomioon, että paikannimiä annetaan usein olemassa olevan nimistön mukaan. Tällöin puhutaan analogiasta ja analogisesta nimenannosta. Esimerkiksi ensimmäisen partisiipin sisältämistä nimistä (esim. *Törisevä, Kohiseva*) lähes puolet on syntynyt olemassa olleiden, samantyyppisten ja samakantaisten nimien mukaan (Kiviniemi 1971; 1977). Alueilta toisille siirtyneet väestöryhmät ovat usein ottaneet mukaansa lähtöseutunsa nimeämismalleja, ja näin on voitu tutkia asutushistoriaa (esim. Vahtola 1980).

## 3.4. Kontaktionomastiikka

Nimet lainautuvat kielestä toiseen huomattavasti helpommin kuin muut kielenainekset. Lainatut nimet voivat joko muuntua uuteen kieleen sopiviksi tai pysyä alkuperäisessä asussaan. Lisäksi nimet ja nimisysteemit elävät usein rinnan eri kielissä ja kielimuodoissa ja niillä voi olla eri funktioita. Kaikkia näitä kysymyksiä tutkitaan kontaktionomastiikassa.

Suomen kieleen (ja sitä edeltäviin kielimuotoihin) on aina tullut vaikutteita muista kielistä ja kulttuureista. Nämä näkyvät yhä vahvasti nimistössämme. Esimerkiksi kristinusko toi 1100-luvulta alkaen alueellemme kristillisperäistä henkilönnimistöä (esim. *Antti, Lauri*), joka jätti syrjään muinaissuomalaisen nimistön (*Ihamieli, Hyvätoivo*). Tässä muutoksessa oli kyse jo koko nimisysteemin vaihtumisesta toiseen. Nimisysteemin lähes täydellisestä vaihtumisesta tuoreempi esimerkki on puolestaan

Namibian Ambomaasta, jossa suomalaisten ja muiden eurooppalaisten lähetystyö 1900-luvulla aiheutti sen, että monet ambomaalaiset ottivat käyttöönsä eurooppalaisia nimiä (Saarelma-Maunumaa 2003).

Maamme kunnan- ja kylännimistä monet sisältävät jälkiä varhaisista germaanivaikutuksista. Esimerkiksi keskiaikaisen Raaseporin linnaläänin kylännimistä useiden taustalla on germaaninen tai myöhempi skandinaavinen henkilönnimi (Raunamaa 2017). Lisäksi esimerkiksi *Harjavalta*-nimeen sisältyy kantagermaanin kaksiosainen miehennimi **Harja-walða* ja *Hollola*-nimeen germaanisperäinen henkilönnimi *Holl*, *Hollo* tai muu vastaava. Kyseessä ovat siis germaanista lainatut henkilönnimet, joita alueen asukkailla on ollut käytössä. Sen sijaan nimet eivät luonnollisestikaan osoita, että alueella olisi asunut germaanista kieltä puhunutta väestöä. Jotta voidaan päätellä nimien kielellinen alkuperä, tarvitaan pohjaksi leksikaalis-etymologinen analyysi (ks. luku 3.3.). Nimen kielen alkuperän selvittäminen ei aina ole helppoa. Jos se pystytään selvittämään, voidaan usein päätellä jotain myös nimenannon ajasta. Tämä taas johtuu siitä, että kielihistorian, historian ja arkeologian avulla tiedetään, milloin ja missä tiettyjä kieliä on voitu puhua.

Koska Suomessa on jo kauan puhuttu pääosin suomea, ruotsia ja saamen kieliä näiden kielimuotojen edeltäjineen, on näiden nimien ja nimisysteemien kohtaamista eritoten paikannimistössä myös tutkittu ahkerasti. Kieliraja-alueille on syntynyt runsaasti lainanimistöä. Eteläisintäkin Suomea myöten on saamelaisperäisiä paikannimiä: esimerkiksi *Kukasjärvi* sisältää pohjoissaamen 'pitkää' merkitsevän sanan *guhkes*. Parhaiten tunnettuja ovat kuitenkin suomalais-ruotsalaisten alueiden substraattinimistöt eli ne nimet, joita näillä alueilla on käytetty ennen kuin sinne on tullut myöhempää, erikielistä asutusta (esim. *Mossala* < *Mustasalo*) ja jotka sitten uusien asukkaiden suussa ovat mukautuneet lainanottajakielen puhujille paremmin sopiviksi (ks. esim. Pitkänen 1985). *Substraattinimi* määritellään jossakin maassa tai jollakin alueella aiemmin puhuttuun kieleen perustuvaksi nimeksi. *Lainanimi* puolestaan on kielestä toiseen lainattu saman kohteen nimi. Kun nimiä lainataan kielestä toiseen, käännetään nimi vain harvoin. Tavallisempaa on, että nimi lainataan äännemukaelmana: joko kokomukaelmana (*Laajalahti* > *Lalax*, *Bredvik* > *Preiviikki*, *Preiviikin/lahti*) tai osamukaelmana (*Prei/lahti*).

Kaksi- ja monikielisillä alueilla nimet ja nimisysteemit ovat jatkuvasti kosketuksissa keskenään. Suomen virallisesti kaksikielisillä alueilla kaavanimet annetaan molemmilla kielillä. Alueen puhujat voivat käyttää kummankinkielisiä nimiä, ja varsinkin vähemmistökieliin kuuluvat saattavat usein puhua paikoista enemmistökielen nimeä käyttäen. Esimerkiksi Helsingissä ruotsinkieliset voivat paitsi käyttää suomenkielistä nimeä myös muodostaa ruotsinkielisen nimen suomenkielisen pohjalta (esim. *Pengergatan* po. *Terrassgatan*) (ks. Vidberg 2016).

Kaksi- ja useampikielisissä perheissä kielten kohtaaminen näkyy usein konkreettisesti lasten nimenvalinnassa. Nimiksi pyritään antamaan monesti sellaisia nimiä, jotka toimivat molemmissa kielissä. Niitä nimiä taas vältetään, jotka olisivat toiselle kielelle äänteellisesti hankalia tai epämiellyttäviä merkityssisältöjä kantavia.

Menneinä vuosina kielten väliset kontaktit tapahtuivat vain jollain tietyllä maantieteellisellä alueella, mutta internet ja ihmisten ja väestöjen yhä vilkkaampi ja laajempi liikkuvuus ovat muuttaneet kielikontaktit globaaleiksi. Ilmiön vaikutus näkyy paitsi henkilönnimistössä myös erityisesti kaupallisissa nimissä, joiden kielellisen alkuperän analysoiminen avaa kiinnostavia näkökulmia muun muassa retoriikan, kielimaiseman ja kielen valtarakenteiden tutkimiseen (ks. luvut 3.8. ja 3.9.). (Yritysnimien kielellisen alkuperän selvittämisestä ks. Sjöblom 2006.)

## 3.5. Sosiolingvistinen variaationanalyysi

Nimien sosiolingvististä variaatiota eli vaihtelua tutkittaessa kohteena on joko sosiaalinen eli käyttäjäkohtainen tai tilanteinen eli tilannekohtainen variaatio. Sosiaalista variaatiota tutkitaan ottamalla huomioon erilaisia kielenulkoisia muuttujia, kuten ikä, sukupuoli, syntyperä, ammatti tai koulutus ja (muu) sosioekonominen asema. Tässä luvussa keskitytään sosiaaliseen variaatioon, koska tilanteinen variaatio nousee esiin myös luvuissa 3.6. ja 3.7.

Sosiaalista variaatiota on tutkittu ennen kaikkea paikannimistössä. On koottu paikannimiaineistoja haastattelemalla saman alueen asukkaita (ks. Haastattelu aineistonkeruun metodina sosiolingvistiikassa tk.) ja

sitten verrattu näitä nimistöjä toisiinsa. Näistä aineistoista on selvitetty, kuinka paljon ja millaista paikannimistöä eri-ikäiset, eri sukupuolta olevat ja eri ammateissa toimivat omalla asuinseudullaan tuntevat ja käyttävät (esim. Ainiala, Komppa, Mallat & Pitkänen 2000). Erikseen on tutkittu myös lasten nimistönkäyttöä ja havaittu esimerkiksi, että pelkän paikan lajin ilmaisevat appellatiivit riittävät varsinkin pienemmillä lapsilla usein nimiksi (Tikka 2009). Ylipäänsä nimikompetenssia tutkittaessa on lisäksi pyritty löytämään syitä sille, miksi yksilöiden nimivarastot eroavat toisistaan. Vaikka varhaisimmissa tutkimuksissa (esim. Slotte, Zilliacus & Harling 1973) sukupuolta ja ikää pidettiin merkittävämpinä erojen selittäjinä, on myöhempi tutkimus tuonut esiin, että pikemmin eroja selittävät ammatti ja harrastukset (Ainiala, Komppa, Mallat & Pitkänen 2000). Tällaisia tutkimuksia on kutsuttu nimitaidon ja (paikan)nimikompetenssin tutkimuksiksi; edellistä käsitettä voi pitää jo hieman vanhentuneena.

Edellä mainitut nimikompetenssin tutkimukset ovat kohdistuneet maaseudun nimistön tarkasteluun. Kaupunkiympäristö on tyypillisesti monikerroksisempi ja heterogeenisempi, mikä tulee ottaa huomioon tutkimusaineistojen ja -metodienkin valinnassa. Kaupungissa erilaisia sosiaalisia muuttujia on helposti tarjolla enemmän kuin pienemmässä maaseudun yhteisössä. Lisäksi kaupunkilaisen nimistössä virallista ja muutoin suunniteltua nimistöä on suhteessa enemmän kuin maaseudulla. Virallisten ja epävirallisten nimien käytön erot ja erilaiset käyttöfunktiot ovatkin kiinnostava kenttä selvitettäväksi. Slangiasuiset paikannimet voivat olla tyypillisiä enimmälti nuorille, ja niiden avulla voidaan erottautua esimerkiksi vanhemmista ikäryhmistä. Toisaalta joitakin slanginimiä voidaan käyttää hyvin yleisesti. Esimerkiksi *Stadi* ('Helsinki'), *Lintsi* ('Linnanmäki') ja *Aleksi* ('Aleksanterinkatu') tunnetaan laajalti muuallakin kuin Helsingissä, ja ne ovat jopa osa institutionaalista ja kaupallista kielenkäyttöä. Slanginimien käytön ja identiteetin kytköstä tarkastellaan luvussa 3.7.

Sosiolingvistinen variaationanalyysi ei rajoitu vain paikannimistön tutkimukseen, vaikka sitä sillä saralla onkin eniten tehty. Samoin metodein voi tutkia muutakin nimistöä.

## 3.6. Nimet vuorovaikutuksessa

Nimet kuuluvat keskeisenä osana arkiseen ja institutionaaliseen vuorovaikutukseen. Nimiä käytetään keskusteluissa viittaamaan henkilöihin, paikkoihin, tuotteisiin, tapahtumiin ja moniin muihin tarkoitteisiin. Keskustelussa samoihin kohteisiin voidaan viitata lukuisin eri tavoin, eikä viittaus välttämättä aina ole edes propri (esim. *tyttö*, *tuo tyttö*; *kaupunginosa*; vrt. *Maija*; *Sörnäinen*, *Sörkkä*). Keskusteluaineistojen ja keskusteluanalyysin käyttö (ks. Haastattelu aineistonkeruun metodina sosiolingvistiikassa; Keskustelunanalyysi ja vuorovaikutuslingvistiikka tk.) yhdistettynä nimistön tuntemukseen valottaakin proprien ja appellatiivien välistä eroa ja rajapintaa. Tavoitteena on selvittää, millä eri tavoin keskusteluissa yksilöidään ja mitä tehtäviä yksilöivillä ilmauksilla on.

Kun tutkitaan nimiä vuorovaikutuksessa, niitä tarkastellaan kielen aineksina osana kielen järjestelmää ja erityisesti yhteydessä siihen sosiaaliseen toimintaan, jota kielellä luodaan. Esimerkiksi arkikeskusteluissa nimiä käytetään joustavasti muiden kielenainesten lomassa, eikä ole välttämättä mielekästä tarkastella propreja ennalta annettuna kategoriana erillään muista viittaamiseen käytetyistä kielenaineksista varsinkaan, kun ilmauksen proprisuus tai appellatiivisuus selviää usein vasta kontekstista (vrt. luku 1.). Olennaista on viittaaminen, yhteisten referenttien luominen keskusteluun ja ne toiminnot, joiden osana viittaaminen tehdään. Kysymys nimien käytöstä kytkeytyy laajemmin tarkoitteisiin viittaamiseen ja kysymyksiin siitä, millaista tunnistamista ja yhteisyyttä keskustelijat tarvitsevat tai tavoittelevat. Tällaista tutkimusta on tehty suomalaisista keskusteluista ennen muuta pronomini-viittauksista (esim. Seppänen 1998; Etelämäki 2006). Keskustelunanalyyttista tutkimusta henkilöihin ja paikkoihin viittaamisesta nimiä käyttäen on tehty jonkin verran muista kielistä (esim. Enfield & Stivers 2007), mutta suomen kielestä tällaista tutkimusta on toistaiseksi hyvin vähän.

Nimien käyttö ja suhtautuminen nimiin on hyvin kulttuurisidonnaista, ja olemassa oleva tutkimus muista kulttuureista tarjoaa kiinnostavan mahdollisuuden vertailuihin. Nimien käyttö voi esimerkiksi heijastaa valtarakenteita: henkilöitä ei välttämättä kaikissa tilanteissa identifioida heidän omalla nimellään vaan suhteessa muihin ihmisiin (esim. *Matti*

*Virtasen rouva*). Nimien käytöllä voi myös korostaa itseään, liittää arvostetuiksi ajateltuja ihmisiä tai paikkoja omaan lähipiiriinsä (ns. nimien pudottelu).

Nimet voivat joskus nousta keskusteluissa myös puheenaiheeksi, jolloin keskustelijat kommentoivat nimiä ja niihin liittämiään käsityksiä ja merkityksiä. Tällainen metalingvistinen nimien käyttö on myös kiinnostava tutkimuskohde. Varsinkin haastattelupuheessa tällaista puhetta voi olla paljon. Nimiin liitettyjä merkityksiä tarkastellaan erityisesti kansanonomastiikassa.

## 3.7. Kansanonomastiikka

Kun tutkitaan sitä, millaisia käsityksiä ja havaintoja maallikoilla (ei-kielentutkijoilla) on nimistä ja nimistönkäytöstä, tehdään kansanonomastista tutkimusta. Voidaan puhua myös kansanlingvistisestä nimistöntutkimuksesta. Nimiä koskeva puhe kohdistuu joko itse nimiin ja nimielementteihin (metakieli 1) tai nimienkäyttöön ja nimenkäyttäjiin liittyviin uskomuksiin (metakieli 2). Nimenkäyttäjä voi luonnehtia jotakin nimeä (esim. *Juhannusruusunkuja*) vaikka ihanaksi tai suloiseksi (metakieli 1) mutta samalla todeta, että ”sekaisinhan niis ruusuis menee” viitaten siihen, että alueella on muitakin ruusujen mukaan nimettyjä katuja (metakieli 2). Edelliset luonnehdinnat kohdistuvat siis nimen leksikaaliseen tai muuhun kielelliseen sisältöön ja ulkoasuun (nimet sellaisinaan), jälkimmäiset nimen käytettävyyteen (nimet välineinä).

Tyypillistä aineistoa kansanonomastiselle tutkimukselle ovat haastattelu- ja keskusteluaineistot. Keskustelut voivat olla myös verkkokeskusteluja. Kyselylomakkeilla voidaan puolestaan saada nopeasti suurehkojakin aineistoja. Kun kielenkäyttäjät raportoivat omasta tai muiden nimistönkäytöstä, he tulevat samalla myös esittäneeksi käsityksiään siitä, miten he itse tai muut tutkituista kohteista puhuvat. Joskus raportoitu käyttö voi olla ristiriidassa todellisen, spontaanin käytön kanssa. Moni voi esimerkiksi ilmoittaa käyttävänsä slangipaikannimeä (esim. *Stadi*), mutta todellisuudessa virallinen nimi (*Helsinki*) voi olla taajemmin käytössä. Tällöin raportoidulla käytöllä kuvataan ensisijaisesti

identifioitumista jonkin ryhmän jäseneksi tai ilmennetään tiettyä identiteettiä. (Ks. esim. Ainiala & Lappalainen 2010.)

Kansanonomastinen lähestymistapa antaa tutkijalle monipuolista tietoa myös siitä, millaisia mielikuvia ja merkityksiä nimille annetaan ja millaisia asenteita niihin liitetään. Vastaajalla voi olla käytössään valmiita vaihtoehtoja, esimerkiksi erilaisia adjektiiveja (esim. miellyttävä – epämiellyttävä, vanhanaikainen – nuorekas, urbaani – maalainen, maskuliinen – feminiininen), joita hän liittää kysyttyihin nimiin. Verkkokeskustelut taas tarjoavat aineistoja, joissa keskustellaan monesti tunnepitoisestikin esimerkiksi erilaisista etunimistä tai yritysnimistä ja tuodaan esiin näihin nimiin kytkettyjä asenteita ja mielikuvia. Keskustelijoilla voi myös olla omia käsitteitä tai kategorisointeja nimille (esim. ”lähiönimi”, ”mummonimi”). Tutkija pääsee silloin paneutumaan siihen, millaisiin nimiin tällaisia kategorisointeja kohdistetaan ja millä tavoin perustellen. Tutkimus voi näin saada myös diskurssianalyyttisia piirteitä.

## 3.8. Tekstianalyyttinen ja diskurssianalyyttinen metodi

Nimillä on yksilöivän tehtävänsä lisäksi monia muitakin tehtäviä. Nämä voivat liittyä tekstin tehtäviin ja esimerkiksi tukea tekstin lajiominaisuuksia. Nimiä voi teksteissä tutkia esimerkiksi selvittämällä niiden narratiivista funktiota (mikä on nimien osuus tekstin ja tarinan kerronnassa), affektiivista funktiota (miten nimet heijastavat erilaisia tunnetiloja ja luovat tekstin tunneilmastoa) tai ideologista funktiota (miten nimi tukee tai rakentaa tekstin ideologista sanomaa).

Erilaisissa teksteissä ja tekstilajeissa yksilöitäviin kohteisiin voidaan viitata monin eri tavoin. Jos esimerkiksi lehdistötiedotteissa käytetään epävirallisia nimiä (esim. *Kaivari*, *Espa*) virallisten (*Kaivopuisto*, *Esplanadinpuisto*) sijaan, esitetään myös paikkoihin liitettävät tapahtumat epävirallisina ja luodaan läheistä ja tuttavallista vaikutelmaa.

Henkilöihin voidaan teksteissä viitata proprin tai appellatiivin sisältävällä NP:llä. Pelkän etunimen käyttö osoittaa usein tuttavallisuutta ja läheisyyttä. Ei-propriset nimitykset puolestaan ovat tavallisia joissakin uutistyypeissä (onnettomuuksista, rikoksista yms. kertovissa), joissa

viitataan toimijoihin mainitsemalla esimerkiksi henkilön ikä ja sukupuoli (*vuonna 1986 syntynyt mies*).

Varsinkin kaunokirjallisuudessa nimillä on narratiivinen, tarinan kerrontaan liittyvä tehtävä. Nimet ovat vuorovaikutuksessa muiden teoksen sisältöä rakentavien ja tarinaa kuljettavien elementtien kanssa. Nimellä voi kaunokirjallisuudessa olla monia muitakin tehtäviä; nimi voi sisältää esimerkiksi erilaista henkilöä luonnehtivaa tai muuten teoksen sisältöön liittyvää informaatiota. Nimen avulla henkilö voidaan myös kytkeä tiettyyn ajankohtaan tai paikkaan.

Retoriselta kannalta kiinnostavaa on ylipäätään se, mitä nimiä tekstiin on valittu ja miten niitä käytetään, kuinka paljon ja millä tavoin niitä nostetaan esiin ja millaisia ne ovat. Esimerkiksi mainostekstissä tuotteen nimi voi osoittautua keskeiseksi lukijaan vaikuttamisen keinoksi. Muissa tekstilajeissa luonnollistuneista puhetavoista voi kertoa myös esimerkiksi se, käytetäänkö etunimiä, sukunimiä, näiden yhdistelmiä vai epävirallisia nimiä ja mitä määritteitä nimiin lisätään. Diskurssien rakentumista voi siis tutkia myös proprien kautta (vrt. Diskurssianalyysi tk.). Kun tutkitaan nimiä diskursseissa, voidaan kysyä, millaisia ideologisia funktiota nimillä on. Kriittinen onomastiikka pyrkii paljastamaan, miten nimet palvelevat vallankäyttöä. Tällaisiin kysymyksiin pureudutaan usein myös nimimaisematutkimuksessa.

## 3.9. Nimimaisematutkimus ja multimodaalinen analyysi

Tieteenalaa, jossa tutkitaan julkisessa tilassa näkyviä kieliä ja eri kielten suhdetta toisiinsa, kutsutaan *kielimaisematutkimukseksi*. Tällaisen tutkimuksen avulla voidaan selvittää, miten tilaa rakennetaan ja millaisia sosiaalisia ja poliittisia suhteita julkiset tilat ilmentävät. Hyvin suuri osa jokapäiväisessä ympäristössämme näkyvistä kielellisistä ilmauksista on erisnimiä: tienviitoissa ja katukylteissä on paikannimiä, mainostauluissa ja liiketiloissa yritysten ja brändien nimiä, rakennusten seinissä koulujen ja hallintoelinten nimiä. Siksi onomastiikan näkökulmat ovat kielimaisematutkimukselle arvokkaita. Kun keskitytään propreihin, voidaan puhua nimimaisematutkimuksesta. (Ks. myös Puzey 2016; Syrjälä 2018.)

Nimimaisematutkimus vaatii kenttätyötä. Aineisto kerätään valokuvaamalla esimerkiksi jonkin alueen kaikki kadunnimikyltit tai liikehuoneistojen seinillä, ovissa ja ikkunoissa olevat tekstit. Nimimaisematutkijan mielenkiinto kohdistuu usein siihen, millainen asema vähemmistökielillä on julkisessa tilassa. Hän voi esimerkiksi selvittää, kuinka paljon ja mitä gaelinkielisiä nimiä näkyy Skotlannin maanteiden varsilla ja missä järjestyksessä gaelin- ja englanninkieliset nimet esitetään. Todellista tilannetta voidaan sitten verrata siihen, mitä vallanpitäjät väittävät kielipolitiikasta. Näin nimimaisematutkimus on samalla kriittistä onomastiikkaa, jonka tavoitteena on paljastaa piileviä valtarakenteita ja asenteita.

Julkisessa tilassa näkyvät kaupalliset nimet tarjoavat tutkijalle myös muunlaisia näkökulmia. Eri kielten esiintyminen kauppojen nimikylteissä voi kertoa esimerkiksi kaupungin asuinalueen monikielisyydestä, mutta kielivalinnat voivat välittää myös kaupallisia merkityksiä, jotka eivät ole tulkittavissa nimeen sisältyvistä sanoista itsestään: ranskankielisillä nimillä rakennetaan erilaista mielikuvaa kuin italian- tai englanninkielisillä nimillä. Kaupallisille nimille kielen elementit ovat yleensä vain yksi mielikuvia välittävä keino. Nimiin sisältyvien sanojen ja muiden kielen elementtien lisäksi niiden merkitykset muodostuvat myös muista semioottisista resursseista, kuten nimen koosta, kirjainten muodosta ja väristä tai nimen liikkumisesta. Kaupallisen nimimaiseman tutkimisessa on mahdollista hyödyntää multimodaalista analyysia (esim. Kress & van Leeuwen 2001).

## 3.10. Suuret korpukset ja niiden määrällinen analyysi

Nimistöntutkijalla on pääsy moniin laajoihin korpuksiin. Kaikkien suomalaisten viralliset henkilönnimet ovat käytettävissä Väestörekisterikeskuksessa, ja mittava, 2,7 miljoonan nimen kokoelma perinnäisiä paikannimiä on puolestaan digitaalisessa Nimiarkistossa sekä Maanmittauslaitoksen Paikannimirekisterin ja Nimiarkiston aineiston sisältävässä Nimisammossa[4]. Rekisteröidyt kaupalliset nimet ovat saatavissa Patentti- ja rekisterihallituksesta, ja esimerkiksi monia eläinten (hevoset,

koirat ym.) virallisia nimiä on omissa rekistereissään. Internetin nimistöä on niin ikään suurina korpuksina tutkittavaksi.

Henkilönnimikorpukset tarjoavat mahdollisuuden tutkia muiden muassa nimien suosionvaihtelua ja eri nimien ja nimityyppien jakautumista. Etunimien suosionvaihtelua on tutkittu, kuten muidenkin kulttuuristen innovaatioiden leviämistä, käyttäen hyväksi E. M. Rogersin luomaa diffuusiotutkimuksen käsitettä ja metodia. Siinä uusien ilmiöiden leviämistä tarkastellaan sekä sosiologiselta että maantieteelliseltä kannalta. Etunimistön osalta on saatu selville, että pääosa nimistä noudattaa leviämisessään samankaltaista suosiokäyrää, jossa nimi nousee määrällisesti suosionsa huipulle ja sitten suosio laskee. Yleensä nimi tulee uudelleen suosituksi noin sata vuotta ensimmäisen suosionsa jälkeen. Uusien nimien suosio nousee usein jyrkemmin ja korkeammalle kuin perinteisten nimien. (Ks. esim. Kiviniemi 2006.)

Suomalaisten sukunimet sisältyvät niin ikään Väestörekisterikeskuksen aineistoihin. Samalla tavoin kuin etunimistä myös sukunimistä voidaan tutkia eri nimien yleisyyttä ja esimerkiksi alueellista jakautumista. Myös eri nimityyppien (kantasanatyyppiset, yhdyssanatyyppiset ja johtimelliset sukunimet) osuuksia eri aikakausina voidaan tutkia. On saatu selville muun muassa, että johtimellisia nimiä on 40 % kaikista suomalaisista sukunimistä, mutta niiden kantajia on jopa 62 % suomalaisten sukunimien kantajista.

Digitaalinen nimiarkisto sisältää noin 2,7 miljoonaa Suomesta ja Suomen lähialueilta kerättyä paikannimeä, joihin sisältyy myös karttatieto, eli käyttäjä näkee kartalta, missä paikat sijaitsevat. Aineiston avulla voidaan tutkia esimerkiksi eri nimien ja nimielementtien esiintymistä ja levikkejä. Esiin saadaan tietoja vaikkapa siitä, missä ja millaisissa nimissä esiintyy perusosana *-töyry* tai *-stO*-suffiksi.

Internetin erilaiset nimistöt tarjoavat monipuolisia tutkimusmahdollisuuksia. Tilastollisia menetelmiä apuna käyttäen voidaan tutkia vaikkapa käyttäjänimien yleisiä piirteitä eri verkkoyhteisöissä. Tiedonlouhintaa, todennäköisyyslaskentaa ja tilastollisia analyyseja on käyttänyt metodeinaan esimerkiksi Antti Leino (2007) tutkiessaan Maanmittauslaitoksen paikannimirekisteristä poimimiensa 58 000 suomalaisen järvennimen perusteella nimeämismalleja ja nimianalogioita.

# 4. Yhteenveto

Proprit on sanaluokka, joka sisältää rakenteeltaan ja semanttiselta sisällöltään monenkirjavan joukon erilaisiin tarkoitteisiin viittaavia ilmauksia. Siksi proprien tutkimus ei voi tukeutua vain yhteen metodiin, vaan menetelmä on valittava tutkimuksen kohteen, tutkimusnäkökulman ja tutkimustavoitteiden mukaisesti. Perinteisesti suomalaisessa onomastiikassa on tutkittu paikannimien typologiaa ja alkuperää ja tutkimukseen on sovellettu syntaktis-semanttista ja leksikaalis-etymologista metodia, mutta viime vuosikymmeninä nimistöntutkijat ovat ryhtyneet tarkastelemaan kohdettaan useammista eri näkökulmista ja yhä enemmän osana koko kielijärjestelmää. Sen vuoksi nimistöntutkimukseen on mahdollista soveltaa monia yleisiä kielitieteen menetelmiä, kuten haastattelututkimuksia, variaatioanalyysia, kansanlingvistiikkaa, vuorovaikutusanalyysia tai kriittistä diskurssianalyysia.

Olemme edellä esittäneet mielestämme keskeisimmät nimistöntutkijoiden nykyään käyttämät metodit. Usein tutkijat myös yhdistelevät eri metodeja saadakseen kattavan kuvan omasta tutkimuskohteestaan. Epäilemättä metodeja on olemassa muitakin, ja uusien kielen teorioiden, tutkimusasetelmien, tutkimuskohteiden ja tarkastelukulmien myötä tutkijoille tarjoutuu yhä uusia tapoja proprien tarkasteluun. Luultavasti onomastiikka kytkeytyy tulevaisuudessa yhä kiinteämmin myös kieliopin tutkimukseen ja tuo näin proprit muiden sanojen rinnalle tasavertaiseksi osaksi kielijärjestelmää.

**Aiheesta lisää:**

Ainiala, Terhi, Saarelma, Minna & Sjöblom, Paula. 2008. *Nimistöntutkimuksen perusteet*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Ainiala, Terhi, Saarelma, Minna & Sjöblom, Paula. 2012. *Names in Focus. An Introduction to Finnish Onomastics*. Helsinki: Finnish Literature Society.

Hough, Carole & Izdebska, Daria (toim.). 2016. *The Oxford Handbook of Names and Naming*. Oxford: Oxford University Press.

Paikkala, Sirkka (päätoim.). 2007. *Suomalainen paikannimikirja*. Helsinki: Kotimaisten kielten tutkimuskeskus ja Affecto Finland.

Julkaisusarjat: *Names*, *Namn och bygd*, *Norna-rapporter*, *Onoma*, *Studia anthroponymica Scandinavica*.

## VIITTEET

1 Suomalaisten hevosten nimien ominaispiirteitä on tutkinut Marja Kalske (2005).
2 Lastenkirjallisuuden nimistä on Suomessa kirjoitettu yksi väitöskirja (Bertills 2003).
3 Kognitiiviseen kielikäsitykseen perustuu osittain myös Willy Van Langendonckin (2007) teoreettis-typologinen kokonaiskuvaus.
4 http://nimisampo.fi/app.

## KIRJALLISUUS


Ainiala, Terhi, Komppa, Johanna, Mallat, Kaija & Pitkänen, Ritva Liisa. 2000. Paikannimien käyttö ja osaaminen: nimitaito Pälkäneen Laitikkalassa. *Virittäjä* 104:3, 330–354. Saatavissa: https://journal.fi/virittaja/article/view/40020.

Ainiala, Terhi, Saarelma, Minna & Sjöblom, Paula. 2008. *Nimistöntutkimuksen perusteet*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Ainiala, Terhi & Lappalainen, Hanna. 2010. Miten Helsingistä puhutaan. *Virittäjä* 104:1, 71–107. Saatavissa: https://journal.fi/virittaja/article/view/4312.

Bertills, Yvonne. 2003. *Beyond Identification. Proper Names in Children's Literature*. Åbo: Åbo Akademis förlag.

Enfield, N. J. & Stivers, Tanya. 2007. *Person Reference in Interaction. Linguistic, Cultural and Social Perspectives*. Cambridge: Cambridge University Press.

Etelämäki, Marja. 2006. *Toiminta ja tarkoite. Tutkimus suomen pronominista* tämä. Helsinki: Suomalaisen Kirjallisuuden Seura.

Forsman, A. V. 1891. *Pakanuudenaikainen nimistö. Johdanto. Tutkimuksia Suomenkansan persoonallisen nimistön alalla I*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 25–50.

Gardiner, Alan H. 1940. *The Theory of Proper Names. A Controversial Essay*. London: Oxford University Press.

Hough, Carole & Izdebska, Daria (toim.). 2016. *The Oxford Handbook of Names and Naming*. Oxford: Oxford University Press.
Hämäläinen, Lasse. 2016. Suomalaisten verkkoyhteisöjen käyttäjänimet. *Virittäjä* 120:3, 398–422. Saatavissa: https://journal.fi/virittaja/article/view/51360.
ISK = Hakulinen, Auli, Vilkuna, Maria, Korhonen, Riitta, Koivisto, Vesa, Heinonen, Tarja Riitta & Alho, Irja. 2004. *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Jespersen, Otto. 1958 [1924]. *The Philosophy of Grammar*. (8. p.). London: Allen & Unwin.
Kalske, Marja. 2005. *Suomessa syntyneiden hevosten nimistö*. Turku: Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 73.
Kiviniemi, Eero. 1971. *Suomen partisiippinimistöä*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Kiviniemi, Eero. 1975. *Paikannimien rakennetyypeistä*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Kiviniemi, Eero. 1977. *Väärät vedet. Tutkimus mallien osuudesta nimenmuodostuksessa*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Kiviniemi, Eero. 1987. Voidaanko nimeämisperusteet luokitella? *Kieli* 2, 7–34.
Kiviniemi, Eero. 1990. *Perustietoa paikannimistä*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Kiviniemi, Eero. 2006. *Suomalaisten etunimet*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Kress, Gunther R. & Leeuwen, Theo van. 2001. *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Arnold.
Kripke, Saul A. 1972. Naming and necessity. *Semantics of Natural Language*, 253–355.
Laine, Päivi. 2007. *Suomi tiellä sivistyskieleksi: suomenkielisen maantieteen sanaston kehittyminen ja kehittäminen 1800-luvulla*. Turku: Turun yliopisto.
Langendonck, Willy Van. 2007. *Theory and Typology of Proper Names*. Trends in Linguistics: Studies and monographs 168. Berlin: Mouton de Gruyter.
Leino, Antti. 2007. *On Toponymic Constructions as an Alternative to Naming Patterns in Describing Finnish Lake Names*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Mallat, Kaija. 2007. *Naiset rajalla.* Kyöpeli, Nainen, Naara(s), Neitsyt, Morsian, Akka *ja* Ämmä *Suomen paikannimissä*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Mill, John Stuart. 1906. *A System of Logic: Ratiocinative and Inductive. Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. London: Longmans, Green & co.
Nissilä, Viljo. 1939. *Vuoksen paikannimistö I*. Suomalais-Ugrilaisen Seuran Toimituksia 214. Helsinki: Suomalais-Ugrilainen Seura.
Nissilä, Viljo. 1962. *Suomalaista nimistöntutkimusta*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Nissilä, Viljo. 1975. *Suomen Karjalan nimistö*. Joensuu: Karjalaisen kulttuurin edistämissäätiö.
Paikkala, Sirkka. 2004. *Se tavallinen Virtanen: suomalaisen sukunimikäytännön modernisoituminen*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Paikkala, Sirkka (päätoim.). 2007. *Suomalainen paikannimikirja*. Helsinki: Kotimaisten kielten tutkimuskeskus ja Affecto Finland.
Pitkänen, Ritva Liisa. 1985. *Turunmaan saariston suomalainen lainanimistö*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Puzey, Guy. 2016. Linguistic landscapes. Julkaisussa: Hough, Carole & Izdebska, Daria (toim.) *The Oxford Handbook of Names and Naming*. Oxford: Oxford University Press, 395–411.

Raunamaa, Jaakko. 2017. Finnish medieval village names based on anthroponyms in the Castle Province of Raseborg. Julkaisussa: Ainiala, Terhi & Saarikivi, Janne (toim.) *Personal Name Systems in Finnic and Beyond*. Uralica Helsingiensia 12. Helsinki: Finno-Ugrian Society, 33–82.

Russell, Bertrand. 1956. Lectures on logical atomism. Julkaisussa: Marsh, Robert C. (toim.) *Logic and Knowledge*. London: Georg Allen & Unwin.

Saarelma, Minna. 2006. *Nimipäiväjuhlat*. Helsinki: Kirjapaja.

Saarelma-Maunumaa, Minna. 2003. *Edhina ekogidho – Names as Links: the Encounter between African and European Anthroponymic Systems among the Ambo People in Namibia*. Helsinki: Finnish Literature Society.

Sartjärvi, Maija. 2013. Kahden etunimityypin yleistyminen ja suosionvaihtelu. *Virittäjä* 117:1, 44–69. Saatavissa: https://journal.fi/virittaja/article/view/7844.

Schybergson, Anita. 2009. *Kognitiva system i namngivningen av finländska handelsfartyg 1838–1938*. Nordica Helsingiensia 17. Väitöskirja. Helsingfors: Helsingfors universitet. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-5692-5.

Seppälä, Janne. 2013. *Kulkine.net – tutkimus epävirallisista kulkineennimistä*. Väitöskirja. Jyväskylä: Jyväskylän yliopisto. Saatavissa: http://www.kulkine.net/.

Seppänen, Eeva-Leena. 1998. *Läsnäolon pronominit.* Tämä, tuo, se *ja* hän *viittaamassa keskustelun osallistujaan*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Sjöblom, Paula. 2004. Kognitiivinen näkökulma proprien semantiikkaan. *Virittäjä* 108:1, 80–91. Saatavissa: https://journal.fi/virittaja/article/view/40315.

Sjöblom, Paula. 2006. *Toiminimen toimenkuva: suomalaisen yritysnimistön rakenne ja funktiot*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Slotte, Peter, Zilliacus, Kurt & Harling, Gunilla. 1973. Sociologiska namnstudier. Julkaisussa: Zilliacus, Kurt (toim.) *Synvinklar på ortnamn*. Helsingfors: Svenska Litteratursällskapet i Finland, 97–180.

Syrjälä, Väinö. 2018. *Namn – språk – ställe. Språkbrukarna i Svenskfinlands offentliga rum*. Nordica Helsingiensia 52. Väitöskirja. Helsingfors: Helsingfors universitet. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-4299-3.

Tieteen termipankki. 2019. [verkkoaineisto]. [viitattu 7.2.2019]. Saatavissa: https://tieteentermipankki.fi.

Tikka, Kaisa. 2009. Kontakter inom barnkultur – barn som namngivare och namnbrukare. Julkaisussa: Mattfolk, Leila & Ainiala, Terhi (toim.) *Namn och kulturella kontakter. Handlingar från NORNA:s 37 symposium i Hapsal 22–25 maj 2008*, 73–84. NORNA-rapporter 85. Helsingfors: Forskningscentralen för de inhemska språken.

Vahtola, Jouko. 1980. *Tornionjoki- ja Kemijokilaakson asutuksen synty. Nimistötieteellinen ja historiallinen tutkimus*. Rovaniemi: Pohjois-Suomen historiallinen yhdistys.

Vidberg, Maria. 2016. *Ortnamn i kontakt i Helsingfors. Finska inslag i bruket av gatunamn i svenska kontexter*. Nordica Helsingiensia 45. Väitöskirja. Helsingfors: Helsingfors universitet. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-2100-4.

Virtaranta, Pertti. 1973. Nimistöntutkimuksemme uranuurtaja Viljo Nissilä 70-vuotias. Julkaisussa: *Viljo Nissilälle 21.8.1973*. Castrenianumin toimitteita 6. Helsinki: Suomalais-Ugrilainen Seura, 11–23.
Zilliacus, Kurt. 1966. *Ortnamnen i Houtskär*. Helsingfors: Svenska litteratursällskapet i Finland.

# Luku 20 Tekstintutkimus

*Mikko T. Virtanen*
https://orcid.org/0000-0002-8126-5929
*Pirjo Hiidenmaa*
https://orcid.org/0000-0002-1925-8445

**Mikä?**

Tekstintutkimus hakee systemaattista ymmärrystä siitä, millaisia kirjoitetut tekstit ovat ja miksi ne ovat sellaisia kuin ovat. Se tuottaa tietoa sanaston ja kieliopin käytöstä teksteissä, tekstien rakenteista ja jäsentelytekniikoista sekä keinoista tuottaa tilanteista vuorovaikutusta ja muodostaa yhteyksiä paitsi toisiin teksteihin myös yhteiskuntaan ja kulttuuriin laajemmin. Tekstejä tutkitaan sekä tuotoksina että prosesseina huomioiden toisaalta staattiset piirteet (esim. kokonaisjäsentely), toisaalta dynaamiset, tekstin edetessä kehittyvät piirteet (esim. teeman- ja informaationkulku). Tekstien tutkiminen on tärkeää, sillä jälkiteolliset yhteiskunnat toimivat pitkälti kirjoitettujen tekstien varassa. Tekstintutkimuksen keskeinen menetelmä on tekstianalyysi. Lisäksi voidaan hyödyntää esimerkiksi haastattelu- ja havainnointimenetelmiä.

**Katso myös:**

Luku 5 Lingvistinen etnografia toimintayhteisöissä

Luku 6 Diskurssianalyysi

Luku 22 Keskustelunanalyysi ja vuorovaikutuslingvistiikka

# 1. Johdanto

Tekstintutkimus hakee systemaattista ymmärrystä siitä, millaisia etenkin kirjoitetut tekstit ovat ja miksi ne ovat sellaisia kuin ovat. Vaikka kirjoitettuja tekstejä on lukuisia erilaisia, yhdistävä tekijä on esinemäinen ja (enemmän tai vähemmän) staattinen olomuoto. Fysikaalisilta ominaisuuksiltaan kirjoitetut tekstit ovat valoa heijastavia pintoja, jotka sisältävät kirjoitusmerkeiksi tulkittavia muste-, pikseli- tai muita kuvioita ja niiden yhdistelmiä. Tekstin merkitys syntyy, kun lukija yhdistää merkit ja niiden yhdistelminä muodostuvat rakenteet tekstiä ympäröivään kontekstiin. Myös konteksti on osa merkitystä, ja sekin syntyy vasta lukijan tulkinnassa. Ennen kuin merkit tavoittavat tulkitsijan ja saavat tulkinnan, paperilappu tai näyttöpääte ruutunäkymineen ei kuitenkaan ole esinettä kummempi.

Kirjoitettujen tekstien kulttuurisesti ja historiallisesti merkittävinä ominaisuuksina voi mainita ainakin esinemäisyyden, säilyvyyden, siirrettävyyden, monistettavuuden ja näkyvyyden. Tekstien tutkiminen on tärkeää, sillä jälkiteolliset yhteiskunnat toimivat pitkälti kirjoitettujen tekstien varassa: teksteillä muun muassa dokumentoidaan, tehdään sitovia päätöksiä, opetetaan, analysoidaan, vakiinnutetaan käsityksiä, kommunikoidaan ajan ja paikan yli (toisinaan anonyymistikin), mainostetaan sekä vaikutetaan julkisessa tilassa mielipiteisiin ja mielikuviin. Tekstit ja tekstien tuottaminen ovat tulleet osaksi lähes kaikenlaista työtä (Heikkinen, Hiidenmaa & Tiililä 2001; Johansson, Nuolijärvi & Pyykkö 2011; Tiililä & Karvinen 2017). Etenkin asiantuntijatyössä tekstit ovat jo pitkään olleet erottamaton osa itse työtä – sekä lähtökohtana, tuloksena että työskentelymuotona (Hiidenmaa 2001; ks. myös Lehtinen & Pälli 2016; Mikkola 2016). Tietoisten toimien lisäksi yhteisöt tulevat rakentaneeksi teksteillä ajattelutapojaan, asenteitaan ja näkökulmiaan.

Useimmat tekstin merkityksen kannalta kiinnostavat ilmiöt toteutuvat tekstikokonaisuudessa kaikkien tekstissä hyödynnettyjen kielellisten ja tekstuaalisten keinojen yhteisvaikutuksena. Halutessaan tutkija voi poimia lähes loputtomiin kiinnostavia huomioita mistä tekstistä tahansa

ja koota niistä kiehtovan esityksen. Tämä ei kuitenkaan ole tutkimusta. Tieteellisen tutkimuksen tehtävänä on muodostaa systemaattisia kysymyksiä ja perustella niiden mielekkyys aiemman tutkimuksen ja teoreettisten lähtökohtien avulla. Tekstintutkimuksen tavoitteeksi ei riitä tekstin sisällön selostaminen ja esitystavan parafrasoiminen. Aineiston havainnot on kuvattava tarkoituksenmukaisesti valituin tekstianalyyttisin käsittein, ja lopuksi on tulkittava havainnot kontekstissaan. Jokainen analyysi on samalla oppisuunnan teoreettisten lähtökohtien koettelua – vankistamista tai horjuttamista. Tekstintutkimuksen panos kielentutkimukseen on lisätä ymmärrystä varsinkin kirjoitetusta kielestä sosiaalisen kanssakäymisen välineenä.

Laajasti määriteltynä *tekstillä* tarkoitetaan sosiaalista ja semioottista eli sosiosemioottista kokonaisuutta. Tekstin sosiaalisuudella tarkoitetaan yhtäältä näkemystä siitä, että ihmisyhteisöihin ja -kulttuureihin on vakiintunut tunnistettavia (mutta varioivia ja ajan mittaan muuttuvia) esitystapoja, rakennekaavoja ja puhetapoja, joita hyödynnetään sekä tekstien tuottamisessa että tulkinnassa. Toisaalta tarkoitetaan sitä, että jokainen teksti kiinnittyy lukuisiin aikaisempiin teksteihin vastaamalla niihin, kommentoimalla niitä, toistamalla tai kyseenalaistamalla niiden konventioita – ja niin edelleen. Olennaista on, että teksti ymmärretään kielenkäytön, ei kieliopin yksiköksi: teksti on sosiaalisen kanssakäymisen kokonaisuus, joka voi toteutua niin yhtenä virkkeenä (esim. lappuna *Palaan pian!*) kuin tuhatsivuisena romaanina tai päivästä toiseen jatkuvina hallitusneuvotteluina (ks. Halliday & Hasan 1976, 1).

Tekstin semioottisuus tarkoittaa puolestaan sitä, että teksti on vähintään yhteen merkkijärjestelmään tukeutuva ja sen puitteissa tulkittava merkityskokonaisuus. Näin tarkasteltuna teksti voi olla yhtä hyvin puhuttu kuin kirjoitettu kokonaisuus. Se voi olla myös ei-kielellinen kokonaisuus, esimerkiksi liikennemerkki, kuva tai tanssiesitys. Jos teksti yhdistelee eri merkkijärjestelmiä, puhutaan multimodaalisesta tai -semioottisesta tekstistä (esim. Kress 2009; Bateman, Wildfeuer & Hiippala 2017). Teksti voi toteutua yhteistoiminnallisesti vuoropuheluna (esim. kasvokkaiskeskusteluna tai pikaviestittelynä), tai se voi olla alusta loppuun yhden osallistujan tai osallistujatahon itsenäisesti tuottama (esim. uutisjuttu, karhukirje).

Tämän kokoelman tarpeisiin määrittelemme tekstin edellistä hieman suppeammin: tarkoitamme *tekstillä* tyypillistä lineaarisesti etenevää kirjoitettua tekstiä, joka ei toteudu vuoropuheluna, ja huomiomme koskevat ennen muuta tällaisten tekstien tutkimusta.[1] Keskitymme kirjoitettujen tekstien lingvistiseen analyysiin ja sivuutamme multimodaalisuuden. Tiedämme toki, että mikään teksti ei ole monomodaalinen, pelkästään kielellinen olio, vaan jokaisessa tekstissä on tehty vähintäänkin myös tekstin ulkoasua koskevia valintoja: on aseteltu teksti tietyllä tavoin ja tuotettu kirjoitusmerkit tietyn asuisina.

Lingvistisesti tekstintutkimuksen tiedonintressi on kahtalainen: yhtäältä kiinnostus on siinä, miten kieli toimii tekstimerkitysten tuottamisessa, toisaalta siinä, millaisia merkityksiä tietyissä teksteissä ja konteksteissa tuotetaan ja miten (ks. Hiidenmaa 2000, 184–185). Tekstintutkimuksen olennainen menetelmä on laadullinen tekstianalyysi, johon myös kirjoituksemme keskittyy. Täydentäviä menetelmiä ovat esimerkiksi kirjoittajien ja lukijoiden sekä muiden tahojen haastattelut ja havainnoinnit.

Tekstintutkimuksen kanssa osin limittäinen tutkimusala on diskurssintutkimus (engl. *discourse studies*; ks. Diskurssianalyysi tässä kirjassa). Alojen erottaminen toisistaan on jopa jossain määrin väkinäistä, eikä englanninkielisessä tutkimuskirjallisuudessa monestikaan tehdä vastaavaa jakoa. Onkin kiinnostavaa sinänsä, kuinka tekstintutkimus on suomalaisessa ja ehkä yleisemminkin pohjoismaisessa kielentutkimuksessa kehittynyt omaksi tutkimusalakseen (vrt. esim. Ruotsissa *textforskning*). Vaikka sekä tekstintutkimus että diskurssintutkimus tarkastelevat tekstejä ja niiden kontekstuaalisia – tilanteisia, yhteiskunnallisia ja kulttuurisia – merkityksiä, diskurssintutkimuksessa korostuu selvemmin tekstien yhteisöllinen tai yhteiskunnallinen luonne ja siten usein myös vallan ja ideologisten käytänteiden kriittinen analyysi (ks. Fairclough 2010) ja toisinaan myös pyrkimys positiivisen muutoksen aikaansaamiseen (ks. Bartlett 2017). Käytännössä rajat ovat kuitenkin häilyvät, ja suomalaistutkijat käyttävät toisinaan rinnastusta *tekstin- ja diskurssintutkimus*. Diskurssintutkimuksen kohteena voivat olla kirjoitettujen tekstien ohella myös kasvokkaiset keskustelut (ks. Diskurssianalyysi tk.).

Artikkelimme jäsentyy seuraavasti. Ensin luomme katsauksen tekstintutkimuksen historiaan (luku 2) ja esittelemme tekstintutkimuksen metodologisia lähtökohtia yleisellä tasolla (luku 3). Osa-alueina ovat laadullisuus ja määrällisyys, ongelmanasettelu ja aineiston koostaminen. Tämän jälkeen esittelemme analyysimenetelmiä erityyppisissä tutkimusasetelmissa (luku 4). Lopuksi tarkastelemme tekstintutkimuksen nykykenttää ja ajankohtaisia kysymyksiä (luku 5). Tutkimusesimerkeissä olemme painottaneet omalla oppialallamme fennistiikassa eli suomen kielen tutkimuksessa tehtyä tutkimusta, jonka valtavirtaa voi luonnehtia konstruktivistiseksi: kiinnostus on siinä, kuinka kirjoitetuissa teksteissä tuotetaan merkityksiä ja rakennetaan todellisuutta koskevia käsityksiä (konstruktivismista ks. Berger & Luckmann 1994 [1966]; myös Pietikäinen & Mäntynen 2019).[2]

## 2. Tekstintutkimuksen historiaa

Nykyisen tekstintutkimuksen historia ei ole yksinkertainen kirjoitettava, koska kehityslinjoja ja vaikutteiden antajia on useita. Tekstinäkökulmat ovat tulleet vakiintuneeksi osaksi eri kieliaineita eri polkujen kautta. Polut ja lopputulemat ovat erilaisia sen mukaan, miten tekstintutkijat ovat yhdistelleet retoriikan, filologian ja strukturalistisen kielentutkimuksen perinnettä (ks. myös Kalliokoski 1996a, 24–30).

Koko ihmistieteiden kentän kannalta merkittävää oli se, että 1960-luvulla Berger ja Luckmann (1994 [1966]) muotoilivat teorian todellisuuden sosiaalisesta rakentumisesta. Tämä sai monen tutkijan kiinnostumaan kielestä ja teksteistä. Richard Rorty (1967) alkoikin kutsua tätä suuntausta *kielelliseksi käänteeksi*. Ajatus siis oli, että kielenkäyttö muovaa käsitystämme todellisuudesta ja siten kielen tarkastelu on avain todellisuuden tutkimukseen (ks. myös Kuortti, Mäntynen & Pietikäinen 2008). Sitä mukaa kuin kielentutkijat ovat laventaneet tarkastelua perinteisen kielentutkimuksen ydinalueelta eli äänne-, muoto- ja lauseopin sekä sanaston tarkastelusta merkitysten, ideologioiden, kontekstien

ja kulttuuristen käytänteiden tarkasteluun, keskustelukumppaneiksi on tullut esimerkiksi kulttuurintutkijoita, kasvatustieteilijöitä, antropologeja sekä yhteiskunnan tutkijoita. Näistä jokaiselle teksti on hieman omanlaisensa olio. Sen ohella, että lingvistit hakevat vastauksia tekstin ja tekstiyden olemukseen, etenkin kielentutkimuksen ulkopuolella kiinnostuksen kohteena on ollut tekstien kuljettama merkitys sisällön, erityisesti foucault'laisten diskurssien näkökulmasta (ks. Foucault 2005 [1969]).[3]

Kielentutkijoiden kiinnostus diskurssiin, kielen käyttöön, syntyi monessa vaiheessa, joista voi mainita muun muassa seuraavat. De Saussure (2014 [1916]) erotti kielen ja sen käytön (ra. *langue* ja *parole*). Firth (1957 [1934–1950]) vakiinnutti käsitteen *tilannekonteksti* ja ohjasi etsimään merkitystä kielen tilannesidonnaisesta käytöstä. Myös Austin (2016 [1962]) ja Searle (1969) nostivat tarkasteluun kielellä toimimisen: kuinka kieltä käytetään asioiden tekemiseen. Labov ja Waletzky (1967) puolestaan kuvasivat suullisten henkiinjäämiskertomusten rakenteellisia ja kielellisiä konventioita ja inspiroivat monia tulevia tekstilajitutkijoita. Hallidayn systeemis-funktionaalinen teoria oli ensimmäisiä täysimittaisia teorioita kielen yhteisöllisestä ja vuorovaikutteisesta luonteesta. Erityisesti hänen teoksensa *Language as social semiotic* (1978) kehitteli Firthin ajatuksia kielen ja kontekstin suhteesta eteenpäin. Samoihin aikoihin myös Hymes (1972) siirsi huomion kieliopin säännöistä kielenkäytön sääntöihin toteamalla, että edelliset olisivat hyödyttömiä ilman jälkimmäisiä. Vaikka painotukset ovat muokkautuneet vuosikymmenten aikana, käsitteitä *tilanne, konteksti, funktio, sosiaalinen toiminta, vuorovaikutteisuus, yhteisö, rekisteri* ja *tekstilaji* hyödynnetään tekstintutkimuksessa tänäkin päivänä.

Kielentutkimuksessa oli 1970-luvulle tultaessa syntynyt yhä voimakkaampi kiinnostus kielenkäyttöä sekä kielen sosiaalista ja toiminnallista ulottuvuutta kohtaan. Autonominen ja rakennekeskeinen kielentutkimus sai rinnalleen pragmatiikan, tekstien ja diskurssien tutkimuksen sekä keskustelunanalyysin. Suomessa kielentutkimus oli ollut kauan keskittyneenä äännehistoriaan, sananmuodostukseen ja lauseoppiin. Merkittävä muutos kohti tekstinäkökulmia tapahtui, kun Enkvist käytännössä toi tekstilingvistiikan Suomeen teoksellaan *Tekstilingvistiikan peruskäsitteitä* (1975).

Tässä tekstintutkimuksen varhaisvaiheessa – tekstilingvistiikassa – tutkittiin etenkin tekstinsisäisiä ilmiöitä, kuten koheesiota ja teeman- ja informaationkulkua (ks. luku 4.1.). Lisäksi laadittiin tekstikielioppeja, joilla olisi saatu kieliopin sääntöjen tapaan kuvatuksi, miten tekstit rakentuvat ja etenevät alusta loppuun (ks. van Dijk 1972). Hieman myöhemmin julkaistussa de Beaugranden ja Dresslerin (1981) oppikirjassa *Introduction to text linguistics* käsiteltiin myös tekstuaalisuuden kulttuurisia ja intertekstuaalisia ulottuvuuksia.

Vaikka tekstilingvistiikka oli pitkälti laadullista tutkimusta, Enkvistin johtama tutkimusyksikkö ja varsinkin sen jatkohanke *Suomen kielen kvantitatiivista tekstisyntaksia* (Hakulinen, Karlsson & Vilkuna 1980) hyödynsivät myös määrällisiä menetelmiä tekstipiirteiden kuvauksessa. Tuolloin tietokoneiden yleistyminen tarjosi ylipäätään mahdollisuuksia isojen tekstiaineistojen laskennalliseen käsittelyyn, minkä osoitus on myös Saukkosen (1984) kuvaus *Mistä tyyli syntyy?*. Teoksessa esitelty määrällinen tutkimus hyödynsi monimuuttujaisia dimensioita (esim. abstrakti–konkreettinen, objektiivinen–subjektiivinen) tekstien kielellisten piirteiden stilistiseen analyysiin (vrt. Biber 1988).

Varhaisen tekstilingvistiikan klassikoihin lukeutuu myös Werlichin teos *Typologie der Texte* (1975), jossa esitellyt tekstityypit (narratiivinen, deskriptiivinen, ekspositorinen, argumentoiva ja instruktiivinen) ovat yhä alkeiskurssien vakiosisältöä. Werlichin typologia elää käytännön sovelluksena myös Pisa-kokeiden aineistoissa, sillä koetekstit valitaan siten, että ne edustavat mahdollisimman hyvin Werlichin luokkien mukaisia tekstejä. Typologiaa on sovellettu ahkerasti myös fennistiikassa (esim. Hakulinen 1982; Kauppinen & Laurinen 1984; Juvonen 2014).

Tekstilingvistiikalla oli aluksi soveltavan kielentutkimuksen leima: jo Enkvistin (1975) kirjan syntysanat lausuttiin Suomen soveltavan kielitieteen yhdistyksen AFinLAn kokouksessa, ja kirjoittaja itsekin hahmottelee kirjasessaan sovelluksia äidinkielen ja vieraiden kielten opetuksesta tekoälyn tutkimukseen (mts. 119–128). Tekstinäkökulmat saivat heti vauhtia muissakin sovelluksissa. Esimerkiksi Esko Koivusalon ja Liisa Huovinen-Nybergin kielenhuolto-oppaassa *Selkeä virkakieli* (1978) hyödynnettiin teema–reema-rakennetta ja tekstityyppien luokittelua.

1980–1990-luvuilla huomio siirtyi tekstipinnan rakenteellisesta tarkastelusta tekstin merkityksiin ja funktioihin, ja kielenkäytön kontekstit ja instituutiot nousivat keskeisiksi. Tärkeä käänne oli kielen sosiaalisuutta ja merkitysten kontekstuaalisuutta kuvaavien kieliteorioiden esiinmarssi. Systeemis-funktionaalinen kielioppi ja siitä ponnistava tekstin- ja diskurssintutkimus voimistuivat Suomessa sen jälkeen, kun alan maailmankongressi järjestettiin Helsingissä 1989. Myös genretutkimus alkoi yleistyä. Fennistiikassa tutkimuksen kohteita olivat muun muassa koulun ja oppimisen tekstit (Kauppinen & Laurinen 1984; Karvonen 1995) sekä kieli ja valta ja mediatekstien ideologisuus (Kalliokoski 1996b; Heikkinen 1999).

2000-luvulla tekstianalyysia on enenevästi tuettu ja laajennettu tutkimalla myös tekstien tuottamisen ja tulkinnan olosuhteita ja käytänteitä. Tällöin on siirrytty tekstikeskeisestä tutkimuksesta askel lähemmäs kirjoittaja- ja lukijakeskeistä, etnografista tutkimusta (vrt. Lillis 2013; Hyland 2015). Tämä edellyttää osallistujien haastattelemista, kyselyjen tekemistä tai esimerkiksi tilanteiden havainnointia (ks. esim. Makkonen-Craig 2005; Tiililä 2007; Tiittula 2012). Lisäksi tutkijaluentaa teksteistä on toisinaan täydennetty keräämällä oheisaineistoksi tutkittavaa yhteisöä edustavien lukijoiden tulkintoja eli hyödyntämällä niin sanottua reseptioanalyysia (ks. Rahtu 2012; vrt. Visakko 2015; Tainio & Karvonen 2015). Lukemista ja kirjoittamista voidaan lähestyä myös kognitiivisesta ja psykolingvistisestä näkökulmasta (esim. Sanders, Schilperood & Spooren 2001; Kaakinen & Hyönä 2012).

Olemme kuvanneet tekstintutkimuksen historian kehityskulkuna, jossa tutkimuskohde lavenee rakenteista käyttöön, muodosta merkityksiin, merkityksistä toimintaan, teksteistä instituutioihin ja kulttuurisiin käytänteisiin. Tämä kuvaus on luonnollisesti siloteltu ja yksinkertaistettu, mutta tällaisenaankin se vakuuttanee lukijan siitä, että tekstien tutkimuksella on pitkä ja polveileva traditio ja että monien näkökulmien ytimestä löytyy yhtenäinen kiinteä ydin, kieli autenttisissa tilanteissa palvelemassa osallistujien sosiaalisia tavoitteita.

# 3. Tekstintutkimuksen metodisia lähtökohtia

## 3.1. Laadullista vai määrällistä tutkimusta

Menetelmällisesti tekstintutkimuksen ydintä on laadullinen tekstianalyysi, joka nojaa empiriaan eli konkreettisista tekstiesiintymistä tehtyihin havaintoihin erotuksena puhtaasti teoriaa, tutkimuksen lähtökohtia ja käsitteistöä pohtivalle tutkimukselle. Laadulliselle tutkimukselle ominaisesti tekstianalyysin kohteena ovat merkitykset (ks. Alasuutari 2011 [1993]). Merkitysten ajatellaan rakentuvan kontekstuaalisesti, eikä tekstejä tarkastella itsenäisinä tai sulkeisina merkityskokonaisuuksina. Mäntynen on muotoillut asian niin, että ”[t]utkija voi kontekstin avulla ’avata’ tekstin merkityksen; tämän vuoksi kontekstin analysointi on aidosti osa tekstianalyysia” (Mäntynen 2003, 30–31). Tekstintutkijan tehtävänä on tunnistaa tekstin antamat (yleensä hienovaraiset) vihjeet konteksteista ja perustella, että hänen osoittamansa kontekstit ovat oleellisia analysoitavien merkitysten kannalta.

Konteksteja ei voi analysoida (eikä edes luetella) tyhjentävästi, vaan ne on huomioitava suhteessa tutkimuksen kysymyksenasetteluun. Potentiaalisia, analyysissa mahdollisesti olennaisia konteksteja ovat lokaalit eli tilanteiset kontekstit sekä globaalit eli kulttuuriset, yhteiskunnalliset ja historialliset kontekstit. Kontekstien ei ymmärretä olevan olemassa tekstistä irrallaan, vaan suhde nähdään kaksisuuntaiseksi: teksti paitsi nojaa konteksteihin samanaikaisesti myös rakentaa kuvaa siitä, millainen konteksti on. Riittävä kiinnittyminen globaaleihin konteksteihin on ehto sujuvalle ja ymmärrettävälle kommunikaatiolle etenkin julkisissa ja heterogeeniselle yleisölle suunnatuissa teksteissä (esim. media- ja viranomaistekstit).

Tekstintutkimuksen kontekstuaalisesta analyysiotteesta seuraa, että tekstimerkitysten ei ymmärretä rakentuvan niinkään yksittäisissä sanoissa ja rakenteissa vaan siinä, miten ne yhdessä kontekstoivat toisiaan: vahvistavat tai kumoavat toistensa tulkintoja ja yhteispelinään muodostavat tilanteisesti ja kulttuurisesti ymmärrettäviä merkityksiä ja toimintajaksoja. On myös hyvä pitää mielessä, että tekstiyhteyden eli kotekstin (engl. *co-text*) näkökulmasta tekstin merkitykset ovat

dynaamisia eli kehittyviä ja muuttuvia: tulkinta edeltävästä tekstinosasta toimii kontekstina paraikaa tulkittavalle tekstinosalle, joka puolestaan rekontekstualisoi edeltävää joko vahvistamalla tai horjuttamalla sitä koskevaa aiempaa ymmärrystä (vrt. Keskustelunanalyysi ja vuorovaikutuslingvistiikka tk.). Samoin tekstin lajitulkinta voi muuttua lukemisen edetessä. Lisäksi genrejen ja kulttuuristen diskurssien eli puhetapojen ”näkyminen” tutkittavassa tekstissä on tietenkin suhteellista: jos tutkija ei tunne tiettyä genreä tai diskurssia, hän ei voi myöskään havaita sen toteutumista kielenkäytössä.

Laadullista tekstianalyysia on mahdollista ja melko tavanomaistakin tukea määrällisen analyysin avulla eli laskemalla kielenpiirteiden esiintymämääriä ja vertailemalla niitä tekstinosittain tai suhteessa muun tyyppisiin teksteihin (esim. Kankaanpää 2006; Tiililä 2007). Itse asiassa voi olla vaikea löytää tekstintutkimusta, jossa ei aineistoa kuvattaessa käytettäisi lainkaan suuntaa antavia kvantifioivia ilmauksia, kuten *yleensä*, *usein*, *tavallisesti* tai *harvoin*. Tällaiset luonnehdinnat jäävät kuitenkin kvasikvantifioiviksi, ellei esiintymien määriä osoiteta kurinalaisesti kuvaamalla, mistä otoksesta ja millaisin laskelmin tällaisiin tulkintoihin on päästy. (Heikkinen, Lounela & Voutilainen 2012, 330.) Mikäli tekstijoukkojen määrällinen vertailu on tutkimuksen pääasiallinen menetelmä, tutkimus on todennäköisesti lähempänä korpuslingvistiikkaa kuin tekstintutkimusta (ks. esim. Historiallinen korpuslingvistiikka; Määrällinen korpuslingvistiikka tk.). Tutkimussuuntausten väliset rajat ovat toki häilyviä, ja samassa tutkimuksessa voidaan hyödyntää useita menetelmiä.

## 3.2. Tutkimuskysymyksen muotoilu

Kuten tieteellisessä tutkimuksessa yleensä myös tekstintutkimuksessa prosessi lähtee liikkeelle ongelmanasettelusta: on päätettävä, mitä halutaan tutkia. Syyllistymättä liikaan psykologisointiin voinemme sanoa, että tutkimusongelman muotoutuminen tutkijan päässä on seurausta ainakin kahdesta seikasta: aiemman tutkimuskirjallisuuden tuntemuksesta ja omista empiirisistä tai metateoreettisista havainnoista. Näiden

avulla tutkija perustelee ns. tutkimusaukon, jota hän lähtee omassa työssään täyttämään. Tutkija on voinut esimerkiksi havaita teksteistä ilmiöitä, jotka sotivat tutkimusalalla vallitsevaa ymmärrystä vastaan tai jotka eivät ole selitettävissä aiemman kirjallisuuden tarjoamin teoreettisin välinein taikka joita ei ylipäätään ole tutkittu.

Tekstintutkimuksen peruskysymyksiä ovat esimerkiksi seuraavat:

- Millaisissa tekstiympäristöissä tietynlaisia kielellisiä rakenteita tai sanoja käytetään? Mitä tehtäviä ne näissä ympäristöissä toteuttavat?
- Miten tietynlaisissa tekstiympäristöissä ilmaistaan tietynlaisia merkityksiä?
- Miten tekstit toiminnallisina kokonaisuuksina rakentuvat?
- Kuinka teksteissä suuntaudutaan oletettuun lukijaan tai yleisöön?
- Miten teksteissä rakennetaan yhteyksiä paitsi yksittäisiin toisiin teksteihin myös yhteiskuntaan ja kulttuuriin laajemmin?
- Kuinka tietynlaiset tai tietynlajiset tekstit ovat muuttuneet tietyllä ajanjaksolla?
- Kuinka tekstejä tuotetaan ja tulkitaan?

Kuvauksen (”millaisia”, ”miten”) lisäksi pyritään yleensä myös havaintojen selittämiseen (”miksi”, ”mistä johtuu”).

Tutkimusongelman muotoilu vaikuttaa siihen, millaisia ilmiöitä aineistosta nousee esiin tutkimuksen tuloksina. Hiidenmaa (2000, 185) on havainnollistanut asiaa seuraavasti: ”Tutkimus lehtitekstin olemuksesta nyt ja sata vuotta sitten on erilainen tutkimus, jos kysytään, miten lehtiteksti on muuttunut sadassa vuodessa. Jos taas kysytään, millä tavoin lehtiteksti on erottunut muista teksteistä eri aikoina, löydetään erilaisia piirteitä, vaikka aineisto olisi sama.”

Haastavinta tekstianalyysissa lienee irtautua omista ennakko-oletuksista ja arkiymmärrykseen nojaavasta tekstintulkinnasta – tavoitteena on tekstin analysoiminen mahdollisimman puhtaalta pöydältä. Esimerkiksi arvottavasta evaluoinnista kiinnostunut aloitteleva tutkija saattaa arkiymmärryksen pohjalta olettaa, että kirja-arvioiden kaltaisissa teksteissä evaluointia ilmaistaan ennen muuta ns. asenneadjektiivein

(esim. *hyvä*, *huono*). Jos sitten jo ennen aineistoon tutustumista rajaa evaluoinnin tarkastelun tällaisiin adjektiiveihin, voi tuloksena olla hyvinkin vinoutunut kuvaus siitä, millaisesta toiminnosta tai toiminnasta evaluoinnissa on kyse.

## 3.3. Tekstiaineiston laatu ja laajuus

Tekstintutkijat tarkastelevat yleensä niin sanottuja tosielämän tekstejä, joita voidaan koota tutkimuksen tarpeisiin aidoista esiintymisympäristöistä tai valmiista tekstipankeista (esim. FIN-CLARIN-konsortion Kielipankki[4]). Tietyin reunaehdoin mahdollista on koota aineistoa myös koeasetelmissa. Esimerkiksi koehenkilöiden kirjoitusten avulla on tutkittu kulttuurisia kertomusrakenteita. Kuuluisa on kansainvälinen vertaileva tutkimus, jossa samaa filminpätkää päärynöiden keruusta näytettiin eri maissa ja pyydettiin koehenkilöitä kertomaan, mitä filmissä tapahtuu (Chafe 1980; Kalliokoski 1987).

Aineiston koon ja koostumuksen vaatimukset riippuvat tutkimukselle asetetuista tavoitteista. Tekstintutkijan on ensinnäkin mietittävä, mikä aineiston rooli on tutkimuksen kokonaisuudessa. Aineisto voi olla tutkimuskohde sinänsä. Näin on silloin, kun tutkitaan rajallisesti esimerkiksi tietyn kirjailijan tai yksittäisen teoksen tyyliä (esim. Kalliokoski 2008; Heikkinen & Voutilainen 2009) tai kun etsitään näyttöä oikeudellisia toimia tai rikosten selvittämistä varten (ns. forensinen kielentutkimus, ks. Coulthard & Johnson 2007; Tiililä 2014). Monesti tutkimusorientaatio on kuitenkin tätä laajempi: aineiston katsotaan konstruktivismin hengessä antavan tietoa yhteiskunnassa tai tietyssä yhteisössä vallalla (tai muutoksessa) olevista ajattelu- ja toimintatavoista. Tällöin tutkimuskohteena ei siis ole pelkästään yksittäinen teksti vaan paljon laaja-alaisempi ilmiö, esimerkiksi haastattelupuheeseen perustuvien sitaattien muodostamisen käytänteet journalismissa (Haapanen 2017) tai oppikirjojen tiedonkäsitys (Karvonen 1995). Toki myös jo yksittäisen tekstin analyysi edellyttää sen kiinnittämistä johonkin perinteeseen ja kulttuuriseen kontekstiin.

Tyypillisimmillään aineisto koostuu kokonaisista teksteistä tai tietyn periaatteen mukaan kerätyistä tekstijaksoista (esim. opinnäytteiden

diskussio- eli pohdintaosat; Vuorijärvi 2013). Joillekin kysymyksenasetteluille voi olla tarpeellista ottaa tarkasteluun mukaan myös kuvia ja muita ei-kielellisiä elementtejä. Esimerkiksi jos tutkitaan kuvatekstejä, on tietenkin välttämätöntä huomioida kielellisen ja kuvallisen vaikutus toisiinsa tekstin tulkinnassa (Heikkilä 2006). Multisemioottisuus on huomioitava myös silloin, kun kirjoitettua kieltä on käytetty aineistossa vain yhtenä semioottisena keinona muiden joukossa, kuten monesti mainoksissa ja tuotepakkauksissa.

Aineiston koosta on vaikea antaa yleispäteviä ohjeita. Mikäli tutkimuksen tavoitteena on lähteä liikkeelle tekstiaineistoista ja esittää havaintojen pohjalta yleistyksiä, on todennäköistä, että aineistoksi tarvitaan ainakin kymmeniä, joskus satojakin tekstejä. Jos vaikkapa halutaan muodostaa yleistys siitä, millaisista toiminnallisista vaiheista tai jaksoista työhakemusten perusrakenne koostuu, yhdestä tekstistä saatava evidenssi ei luonnollisestikaan riitä; esimerkiksi Koskelan (2014) tutkimusaineisto koostui 30 työhakemuksesta. Toisaalta jo yksittäinenkin, hyvin valittu tekstiesimerkki voi palvella sen osoittamista, että jokin tutkimusalalle vakiintunut väite tai taustaoletus on liian kapea tai yksioikoinen. Pieni aineisto voi riittää myös silloin, kun otetaan ensi askelia ja kartoitetaan aiemmin tutkimattomia ilmiöitä. Tällaisia ovat 2000-luvulla olleet esimerkiksi tekstien hybridisyyteen eli mm. genrejen yhdistelyyn liittyvät ilmiöt (ks. luku 4.5.). Tekstiaineiston koostamiseen liittyviä kysymyksiä ovat käsitelleet yksityiskohtaisemmin Heikkinen, Lounela ja Voutilainen (2012) sekä Pietikäinen ja Mäntynen (2019, 232–239).

**Mitä tekstintutkija tutkii?**

Tekstintutkimuksen aineistot ovat yleensä aitoja, ilman koejärjestelyitä tuotettuja tekstejä. Kysymyksenasettelut voivat vaihdella pienestä suureen. Voidaan esimerkiksi kysyä, millaisia virkkeenalkuisia kytköksiä tekstissä on ja miten ne rakentavat tekstin merkityksiä. Toisaalta voidaan kysyä, miten uusia tarkoitteita esitellään tekstissä tai miten teksti rakentaa ja heijastaa yhteisön käsityksiä tekstin aiheesta. Koska tekstit ovat yhteisön toiminnan välineitä, tekstien toimintojen ja toimintarakenteiden

kuvaaminen on yksi tekstintutkimuksen tehtävä. Niinpä voidaan kysyä, miten tekstin argumentaatio rakentuu, miten tekstillä ohjaillaan lukijaa ja miten teksti asettuu osaksi yhteisön tekstivarantoa.

# 4. Tekstianalyysin ulottuvuuksia ja näkökulmia

Ongelmanasettelun ja aineistonkeruun jälkeen tekstintutkija voi ryhtyä itse analyysiin. Tämä tarkoittaa aineiston kielellistä lähilukua. Analyysi voi edetä tyvestä latvaan (engl. *bottom up*) eli kielellisestä mikrotasosta kohti tekstin makrotason ilmiöitä, esimerkiksi genreä ja sen mahdollisia toimintavaiheita. Vaihtoehtoisesti voidaan edetä latvasta tyveen (engl. *top down*) eli kokonaisjäsentelystä ja kontekstuaalisista piirteistä kohti tekstin kielellisiä yksityiskohtia. Tutkimuskysymyksestä riippuu, kuinka hienojakoista analyysia on tarpeen tehdä. Tavoitteena ei ole sanoa teksteistä kaikkea vaan tunnistaa ne piirteet, jotka kulloinkin ovat olennaiset selittämään aineiston ominaisuudet ja vastaamaan siitä tehtyyn tutkimuskysymykseen.

Seuraavaksi esittelemme hieman yksityiskohtaisemmin tekstintutkimuksen analyysimenetelmiä erityyppisissä tutkimusasetelmissa. Tarkastelemme ensin koheesion sekä teeman- ja informaationkulun tutkimusta (luku 4.1.), sitten sanaston ja kieliopin käytön tutkimusta (4.2.). Tämän jälkeen käsittelemme diskurssikuvioiden tutkimusta (4.3.), genrejen tutkimusta (4.4.) ja lopuksi moniäänisyyden ja intertekstuaalisuuden tutkimusta (4.5.).

## 4.1. Koheesio sekä teeman- ja informaationkulku

Jo 1970-luvun tekstilingvistiikka oli kiinnostunut tekstinosien yhteenkuuluvuudesta, tekstin etenemisestä ja tutun ja uuden tiedon

vuorottelusta. Niistä ensimmäisellä, *koheesiolla*, tarkoitetaan tekstin muodollista, kieliasusta ilmenevää sidosteisuutta ja yhtenäisyyttä virkkeiden välillä. Koheesiota voidaan tuottaa muun muassa konnektiivein, anaforisin ja kataforisin (eli tekstissä taakse- tai eteenpäin viittaavin) pro-sanoin sekä toiston avulla. Shorea (2012, 182) mukaillen voi todeta, että vaikka lukija ei ymmärtäisi tekstin sisältöä, tällaiset pääasiassa kieliopilliset sidokset ohjaavat häntä tulkitsemaan sisällön suhteita tietyllä tavalla. Astetta implisiittisempiä koheesion keinoja ovat leksikaaliset keinot, esimerkiksi lähimerkityksisten sanojen käyttö (esim. yläkäsite–alakäsite, vastakohtaisuus). Ne ovat implisiittisempiä siksi, että sanojen merkityksiä ja merkityssuhteita koskeva tieto (esim. lekseemien *jäädyke* ja *parfee* välinen synonymia) ei ole oletettavasti kieliyhteisössä yhtä laajasti jaettua kuin kieliopin keinoja koskeva intuitiivinen tieto (ks. Kielitaju kielentutkijan työkaluna tk.). Koheesion tutkimuksen klassikko on Hallidayn ja Hasanin (1976) teos *Cohesion in English*, jota Hakulinen ja Karlsson sovelsivat tuoreeltaan (1979) *Nykysuomen lauseopissaan*. Tekstintutkimuksessa koheesio on noussut keskeiseksi ilmiöksi esimerkiksi oppikirjatekstejä tutkittaessa (Karvonen 1995).

Koheesion lähikäsite on *koherenssi*, jolla tarkoitetaan tekstin funktionaalista ja sisällöllistä järkeenkäypyyttä suhteessa sen sosiokulttuuriseen kontekstiin. *Koherenssin* voi ymmärtää kattokäsitteeksi, jonka alle sisältyy ainakin tekstin genreen, toimintarakenteisiin, moniäänisyyteen ja intertekstuaaliseen liittyvät ilmiöt (ks. myös Rahtu 2006). Niitä tarkastelemme luvuissa 4.3., 4.4. ja 4.5.

Koheesion lisäksi tekstintutkimuksen perinteisiin kohteisiin lukeutuu tekstin teeman- ja informaationkulun tutkimus. Tällöin tutkitaan yhtäältä sitä, millaisin periaattein erityyppisissä teksteissä esitettyä tietoa jaetaan annettuun ja uutisarvoiseen, ja toisaalta sitä, millä tavoin tekstiyhteys vaikuttaa lauseiden jäsennykseen, esimerkiksi sanajärjestykseen tai valinnaisten määriteosien käyttöön (Thompson 2007; Shore 2008). Tällöin tekstiä tarkastellaan prosessina, joka toteutuu vähittäin, lause lauseelta, ja jossa jatkuvasti säädellään tekstiin konstruoitujen kirjoittajan ja lukijan jakaman yhteisen tiedon aluetta (ks. Virtanen, Rahtu & Shore 2018). Prosessi-näkökulmasta tekstin merkitys on siis liikkeessä koko ajan – edeltävässä tekstissä tuotetut merkitykset voivat tulevissa

tekstinosissa tarkentua, laajentua tai vaihtoehtoisesti kyseenalaistua taikka saada uudenlaisen tulkintakehyksen.

## 4.2. Sanaston ja kieliopin käyttö tekstissä

Lingvistisen tekstintutkimuksen peruskysymyksiin on kuulunut se, millaisissa tekstiympäristöissä tietynlaisia kielellisiä rakenteita tai sanoja käytetään ja mitä tehtäviä ne näissä ympäristöissä toteuttavat. Vastakkaiselta suunnalta asetettu mutta yhtä lailla perustava kysymys on, millaisten kielellisten rakenteiden tai sanojen avulla tietynlaisissa tekstiympäristöissä ilmaistaan tietynlaisia merkityksiä. Edellisessä tapauksessa lähdetään siis liikkeelle yksittäisistä kielellisistä resursseista ja selvitetään niiden merkityspotentiaalia tekstimerkitysten rakentamisessa. Tällaisesta lähtökohdasta on tutkittu esimerkiksi journalististen tekstien tietynmuotoisia passiivilauseita (esim. *kysytäänpä asiantuntijalta*), jotka on tutkimuksen tuloksena nimetty dialogisiksi (Makkonen-Craig 2005). Samantapaisesti on tutkittu muun muassa ylioppilasaineiden *että*-lauseyhdistelmiä (esim. *on totta, että*; Juvonen 2010) sekä konditionaalin perfektissä olevia lauseita kirja-arvioissa (esim. *tiivistäminen olisi parantanut teosta*; Virtanen 2015).[5] Lisäksi digitaalisten tekstien tutkimuksessa on tarkasteltu aihetunnisteiden (engl. *hashtag*) ja hyperlinkkien kaltaisia resursseja, joiden analyysissa on huomioitava myös viestintäteknologinen konteksti (ks. esim. Vitikka 2018; Salomaa 2019).

Jos taas kysytään sitä, millaisia merkityksiä teksteissä on ja miten (ja missä määrin) ne syntyvät kielen keinoin, voidaan lähteä liikkeelle esimerkiksi yhdestä toiminnosta ja tarkastella, millaisten kielellisten keinojen avulla se voidaan toteuttaa ja millaiset kontekstuaaliset, esimerkiksi sekventiaaliset eli tekstin jaksotteluun liittyvät seikat keinon valintaan vaikuttavat. Tästä näkökulmasta on tarkasteltu esimerkiksi ohjailua hallinnollisissa kirjeissä (Honkanen 2012) ja mielipidekirjoituksen tekstilajiominaisuuksia sekä tekstilajin ontologiaa ja epistemologiaa yleensä (Nieminen 2010).[6]

Kummassakin edellä esitellyssä lähestymistavassa on etunsa ja rajoitteensa. Tiettyyn kielen keinoon keskittyminen voi tuottaa systemaattisen

analyysin keinon käyttötavoista konteksteineen mutta hataramman kuvan niistä toiminnoista ja toimintajaksoista, joiden toteuttamista keino palvelee. Jos taas asetetaan keskiöön tietty toiminto, voidaan saavuttaa syvällinen ymmärrys itse toiminnosta mutta ei välttämättä niistä kielen keinoista, joiden avulla se toteutetaan. Näin ollen kokonaisvaltaisen ymmärryksen saavuttamiseen siitä, kuinka tekstimerkityksiä tuotetaan ja kuinka tekstuaalinen vuorovaikutus on mahdollista, tarvittaneen sopivassa suhteessa kumpaakin lähestymistapaa. (Vrt. Hakulinen & Selting 2005, 9–10.)

## 4.3. Diskurssikuvioiden tutkimus

Tekstintutkimuksessa on alusta alkaen tutkittu ahkerasti tekstien rakenteita ja jäsentelykaavoja. Rakenne ymmärretään monien muiden kieliopintutkimuksen suuntausten mukaisesti merkityskuvioksi, jonka avulla kielenkäytössä tehdyt yksittäiset valinnat asettuvat osaksi laajempaa, merkitystä kantavaa kokonaisuutta (ks. Shore 2012, 133–134). Siinä missä kieliopintutkimuksen ydintä on syntaktisten rakenteiden tutkimus (ks. Kieliopin tutkimus ja kielioppiteoriat tk.), tekstintutkimuksen kiinnostus on diskursiivisissa rakenteissa eli diskurssikuvioissa, jotka jäsentävät tekstejä virkettä laajemmalla tasolla, esimerkkeinä rakenteet KYSYMYS–VASTAUS, YLEINEN–YKSITYISKOHTAINEN, MYÖNNYTYS–TOIMINTO, TOIMINTO–PERUSTELU.[7] Näistä kahdessa viimeisessä merkintä toiminto tarkoittaa, että rakenneosa voi toteutua lähes millaisena toimintona tahansa, esimerkiksi kysymyksenä, väitteenä tai ehdotuksena (Virtanen 2015, 39–40).

Mainitunlaisista rakenteista on suomalaisessa tekstintutkimuksessa käytetty muun muassa nimityksiä *diskurssikuvio*, *retorinen kuvio* tai *rakenne*, *toimintasarja* sekä *toimintokuvio* (kootusti ks. Virtanen 2015, 40).[8] Tässä käytämme yleiskäsitteenä *diskurssikuviota*. Alan peruskysymys on, millä tavoin tekstit etenevät toiminnosta toiseen, mikä toimintojen keskinäinen työnjako on ja kuinka se käy teksteistä kielellisesti ja diskursiivisesti ilmi. Kuvioita on tutkittu teksteistä eri abstraktiotasoilta, virke- ja kappaletasolta aina lukujen tasolle asti. Lähtökohta onkin ollut,

että diskurssikuviot ovat skaalautuvia, niin että esimerkiksi KYSYMYS–VASTAUS-kuvio voi jäsentää yhtä hyvin kahden virkkeen kuin kahden kappaleen välistä suhdetta (esim. Hoey 2001).

Diskurssikuvioista on julkaistu tutkimuksia ainakin 1970-luvulta alkaen. Suomalaisessakin tekstintutkimuksessa ahkerasti sovellettu malli on William Mannin ja Sandra Thompsonin (1988) kehittämä retorisen rakenteen teoria (engl. *rhetorical structure theory*, RST). Siinä rakenteita kutsutaan retorisiksi siksi, että niiden viimekätiseksi tarkoitusperäksi nähdään lukijan ymmärrykseen ja suhtautumiseen vaikuttaminen. Esimerkiksi VÄITE–PERUSTELU-rakenne luo asetelman, jossa toinen tekstinosa näyttäytyy toisen valossa lukijalle uskottavampana, hyväksyttävämpänä tms. (Mann & Thompson 1988, 251). Teorian kovaa ydintä ovat Mannin ja Thompsonin (mt.) ehdottamat määritelmät noin 20 retoriselle suhteelle. Vaikka tutkija voi halutessaan abstrahoida uusiakin suhdetyyppejä, perusjoukkoon nojaaminen mahdollistaa eri tekstilajeista ja kielialueilta tehtyjen tutkimusten vertailun (Komppa 2012, 180). Vaihtoehtoisia mutta RST:n kanssa pitkälti yhteensopivia malleja ovat tarjonneet muun muassa Eugene Winter (1977; 1994), Michael Hoey (1983; 2001) ja James R. Martin (1992; Martin & Rose 2007 [2003]).

Fennistiikassa diskurssikuvioita on tutkittu esimerkiksi journalistisista teksteistä (Makkonen-Craig 2005), yritysteksteistä (Visakko 2018), lukiolaisten kirjoitelmista (Juvonen 2010; Komppa 2012) ja tieteen teksteistä (Virtanen 2015). Tyypillistä on ollut kiinnostus diskurssikuvioiden avulla rakennettuun vuorovaikutukseen kirjoittajan, lukijan ja mahdollisten kolmansien (referoitujen) osapuolten välillä. Esimerkiksi yritysten visio-, missio- ja strategiatekstejä tutkinut Visakko (2018) on selvittänyt diskurssikuvioiden analyysilla yrityksen ja asiakkaan välille rakennetuissa sosiaalisissa suhteissa esiintyvää vaihtelua. Samanhenkisesti Juvonen (2010) on tarkastellut sitä, kuinka diskurssikuvioita käytetään ylioppilasaineissa oman väitteen asemointiin ja erottamiseen tekstiin tuoduista muista väitteistä.

Diskurssikuviot ovat kielellisen vuorovaikutuksen perusrakenteita: ne eivät ole genresidonnaisia eivätkä myöskään sidottuja kielenkäytön kanavaan – kirjoitetut tekstit ja puhutut keskustelut hyödyntävät samantyyppisiä kuvioita, vaikka tutkimusalojen välinen terminologia

onkin erilaista. Esimerkiksi keskusteluntutkija tarkastelee KYSYMYS–VASTAUS-kuviota kahden osallistujan yhteisesti tuottamana vierusparina (esim. Raevaara 1997), kun taas tekstintutkija tarkastelee sitä kirjoittajan itsenäisesti tuottamana kuviona, jolla on tekstiä jäsentävä tehtävä (esim. Hoey 2001). Lisäksi monet kuvioista lienevät lähes universaaleja, kulttuurista ja kielestä toiseen toistuvia (ks. esim. Enfield, Stivers & Levinson 2010).

## 4.4. Genrejen tutkimus

Diskurssikuvioiden eli tekstien yleisten toimintarakenteiden lisäksi tekstintutkimuksessa on tarkasteltu runsaasti myös yksittäisille tekstilajeille eli genreille luonteenomaisia toimintajaksoja ja kokonaisjäsentelyitä. Tästä näkökulmasta on tutkittu esimerkiksi työhakemuksia (Koskela 2014), tiedotteita (Kankaanpää 2006; Honkanen 2012) sekä monia median genrejä (kootusti ks. Mäntynen 2006). *Genren* käsitteellä viitataan sosiaalisesti vakiintuneisiin toiminnan muotoihin, jotka näkyvät samaa tarkoitusta palvelevien tekstien ainakin jonkinasteisena samankaltaisuutena. Genren tuoma vakioisuus voi koskea tekstirakenteen lisäksi tai asemasta monia muitakin asioita, esimerkiksi sisältöä, vuorovaikutussävyä, taittoa tai julkaisukontekstia. Monesti voidaan myös erottaa genreille ominaisia kieliopillisia ja sanastollisia ”avainpiirteitä” (Heikkinen 2009). Genret ovat olemassa yksittäisten tekstien kautta, ja niiden normit ja konventiot syntyvät, vakiintuvat ja muuttuvat tekstien virrassa.[9] (Ks. myös Diskurssianalyysi tk.)

Genretutkimuksen perusoletuksiin kuuluu, että genret eivät toteudu kielenkäytössä puhtaina vaan kielelliset tuotokset asettuvat eri etäisyyksille genrejen prototyypeistä (esim. Shore & Mäntynen 2006; Martin & Rose 2008; Nieminen 2010). Genret ovat kehikkoja, joiden avulla tekstejä sekä tuotetaan että tulkitaan. Etenkin moniin institutionaalisiin teksteihin voi kyllä liittyä myös pakottavia genrenormeja, jolloin genreä edustavien tekstien välinen variaatio jää melko vähäiseksi. Esimerkiksi kokouksen esityslistan normit edellyttävät määrämuotoisuutta ja lisäksi listan toimittamista kokouskutsun mukana. Päätösesityksessä taas

on ilmaistava tarkasti, onko kyseessä päätös, keskusteluasia vai tiedoksi merkittävä asia. Muilla elämänalueilla tavallisempaa lienee se, että genren normit ovat enemmän tai vähemmän väljät ja joustavat. Siksi menetelmällisesti mielekkäintä onkin, että genretutkija kuvaa ensin aineistonsa yksittäisiä tekstejä erikseen, jolloin tulokset kuvaavat yksittäisiä tekstejä, ja vasta sen jälkeen lähestyy genreä esittämällä siitä tekstianalyyseihin perustuvia yleistyksiä (Hiidenmaa 2000, 185). Bhatia (2004) ja Honkanen (2009) ovat korostaneet myös sen tärkeyttä, että genreanalyysissa huomioitaisiin myös epätyypillisiä ja hankalia tapauksia. Tämä on tärkeää paitsi tutkimuseettisesti myös genreteorian kehittämisen kannalta. Hankalat tapaukset voivat nimittäin paljastaa kysymyksiä tai ulottuvuuksia, joiden suhteen teorian selitysvoima on vielä riittämätön.

Toinen genretutkimuksen perusoletus on, että yhdellä tekstillä voi olla suhde moneen genreen, esimerkiksi niin, että tietty genre toteutetaan tekstissä toisen genren välityksellä. Esimerkiksi mielipidekirjoitus voi toteutua moralisoivan kertomuksen välityksellä ja musiikkiarvio ruokareseptin muodossa (ks. Mäntynen & Shore 2014). Jälkimmäisessä tapauksessa genren valinta on myös osa kannanottoa: reseptin muoto vihjaa, että musiikki on tehty kaavamaisesti. Kun teksti jäsentyy kahden tai useamman genren risteyksessä, voidaan puhua myös hybridisyydestä – ilmiöstä, joka on 2000-luvulla enenevästi kiinnostanut tekstintutkijoita (ks. Visakko & Voutilainen 2012; Mäntynen & Shore 2014; Miller & Bayley 2016).

Tunnettuja genretutkimuksen aloja ovat systeemis-funktionaaliseen perinteeseen nojaava Sydneyn koulukunta (esim. Martin & Rose 2008), englannin ammattikielten tutkimus (*English for Specific Purposes*; Swales 1990; Bhatia 2004) sekä uusretorinen genretutkimus (esim. Miller 1984; Devitt 1991). Alojen välisiä eroja ja yhtäläisyyksiä ovat tarkastelleet muiden muassa Solin (2009) sekä Heikkinen ja Voutilainen (2012). Yhtenä erona on se, missä määrin analyysi kohdistuu teksteihin (tuotoksiin), missä määrin tekstien tuottamiseen ja vastaanottoon sekä yhteisön jäsenten genrekäsityksiin. Esimerkiksi systeemis-funktionaalinen perinne asettaa painon edelliseen, uusretorinen genretutkimus jälkimmäiseen. Erinomainen verkkosivusto genretutkijalle on Genre across borders[10]. Sivusto tuo genretutkimuksen eri perinteitä yhteen ja sisältää muun muassa kattavan termipankin.

## 4.5. Moniäänisyys ja intertekstuaalisuus

Tekstien moniäänisyyden tutkimus pyrkii selvittämään, millä keinoin erityyppisissä teksteissä tuodaan kirjoittajan äänen rinnalle muita ääniä tai näkökulmia ja asetetaan ne vuorovaikutukseen (ks. myös Diskurssianalyysi tk.). Äänen käsite on peräisin Bahtinilta (1981 [1934–1935]), ja se on sukua *intertekstuaalisuuden* eli tekstienvälisyyden käsitteelle, jonka Kristeva (1993 [1967]) esitteli Bahtinin ajatuksiin tukeutuen. Palaamme intertekstuaalisuuteen tuonnempana.

*Moniäänisyydellä* tarkoitetaan vähintään kahden erilaisen äänen tai näkökulman samanaikaista läsnäoloa tekstissä. Konkreettisimmin tämä ilmenee tietenkin referointina, mutta kyse voi olla myös esimerkiksi ironiasta (Rahtu 2006) tai siitä, että tutkittavan tekstin sananvalinnat ja niiden kautta ilmaistut näkökulmat asettuvat vuorovaikutukseen jonkin yhteiskunnallisesti tai kulttuurisesti tunnistettavan ajattelutavan tai käytänteen kanssa. Yksittäisessä tekstissä voi olla kyse vallitsevan ajattelutavan tai käytänteen pönkittämisestä: tekstissä toistetaan (tietoisesti tai tiedostamatta) sitä, mikä on jo entuudestaan vakiintunutta, tai vähintäänkin selitetään, miksi vakiintuneesta poiketaan. Vaihtoehtoisesti teksteissä voidaan asettaa vallitsevia käytänteitä tai puhetapoja kriittisen keskustelun kohteeksi ja pyrkiä uudistamaan niitä.

Vaikka bahtinilaisen ajattelun perusta on, että ollakseen ymmärrettävää kielenkäytön on oltava riittävän moniäänistä – asetuttava riittävän selvästi osaksi aiempaa perinnettä –, tekstintutkimuksessa on siitä huolimatta erotettu omaksi kielenkäytön moodikseen yksiäänisyys. Tällöin *yksiäänisyydellä* tarkoitetaan retorista vaikutelmaa siitä, että tekstissä ei anneta huomiota vastakkaisille tai vaihtoehtoisille näkökulmille (Martin & White 2005, luku 3). Usein kyse on myös siitä, että tekstistä ei käy ilmi, mistä tai keneltä ajatus tai merkitys on peräisin.

Tekstien äänisyyttä on tarkasteltu 2000-luvulla runsaasti systeemis-funktionaalisen suhtautumisen teorian (engl. *appraisal theory*) avulla (ks. Martin & White 2005). Teoria lähestyy tekstejä kirjoittajan hallitsemina dialogisina tiloina (engl. *dialogic space*). Virtanen (2015, 30) on tulkinnut, että *dialogisella tilalla* tarkoitetaan tekstiin konstruoitua vuorovaikutusaluetta, jolla kirjoittaja luo suhteita tekstissä läsnäolevaksi tehtyihin

muihin osapuoliin (ks. Diskurssianalyysi tk.). Vuorovaikutusalue on dynaaminen: sitä säädellään ja muokataan tekstin toiminnosta ja vaiheesta toiseen edetessä. Dialogisen tilan käsitteen avulla voidaan tarkastella esimerkiksi sitä, kuinka paljon oletetulle lukijalle annetaan liikkumavaraa responsiivisena toimijana. Kun dialoginen tila rakennetaan kapeaksi (esimerkiksi kategorisia väitteitä tai käskyjä esittämällä), lukijalta odotetaan samanmielistä tai ainakin myöntyväistä responssia. Kun taas tila konstruoidaan väljäksi (esimerkiksi esittämällä väite varauksellisesti tai käsky pyyntönä), tekstissä osoitetaan eriasteisesti suvaitsevaisuutta myös erimielistä, epäilevää tai vastahankaista lukijaa kohtaan. (Martin & White 2005, luku 3.)

Kuten aiemmin totesimme, tekstien moniäänisyyden tutkiminen tulee lähelle intertekstuaalisuuden tutkimista: molemmissa etsitään merkkejä tai vihjeitä siitä, että teksti on yhteydessä toiseen tekstiin tai yleisemmin esimerkiksi genreen tai diskurssiin. Yhtenä käytännöllisenä erona on kuitenkin se, että *moniäänisyyden* käsitteeseen sisältyvä äänen käsite mahdollistaa *intertekstuaalisuuden* käsitettä paremmin myös tekijän äänen huomioinnin ja sen tarkastelemisen, kuinka tekijän ääni on asemoitu suhteessa tekstiin tuotuihin muihin ääniin (Virtanen 2015, 43).

Yhtä kaikki myös *intertekstuaalisuuden* käsitettä on hyödynnetty tekstintutkimuksessa runsaasti. Tutkimusperinteestä voidaan erottaa ainakin kaksi erilaista suuntausta: yhtäältä tekstistä syntyvien intertekstuaalisten vaikutelmien tutkiminen, toisaalta samaan genreketjuun kuuluvien tekstien vertailu. Siinä missä kieliopintutkijaa kiinnostaa tekstienvälisyyden analyysissa se, millaisin kieliopillisesti kiteytynein keinoin tekstienvälistä suhdetta voidaan ilmaista ja kuinka eri keinot suhteutuvat toisiinsa kielijärjestelmässä (esim. Kuiri 1984; Shore 2005), tekstintutkijaa kiinnostaa ensisijaisesti se, millaisia vakiintuneita käyttötapoja intertekstuaalisuuden ilmenemismuodoilla on eri tekstilajeissa. Voidaan selvittää esimerkiksi sitä, miksi henkilökohtaisissa blogikirjoituksissa referoidaan vuoropuheluja (Virtanen & Juvonen 2018), miksi yleistajuiset kielijutut aloitetaan usein referoinnilla (Mäntynen 2005) tai millä tavoin naispareja referoidaan lehtiartikkeleissa (Uotinen 2007). Tällöin tutkitaan siis intertekstuaalisuuteen liittyviä genrekäytänteitä ja pyritään kirjoittamaan niitä auki.

Metodistisestikin omanlaisekseen tekstienvälisyyden tutkimukseksi hahmottuu tekstilajiketjujen ja rekontekstualisoinnin tutkimus. Tällöin kiinnostuksen kohteena on se, millaisia vakiintuneita, ennakoitavia yhteyksiä kahteen tai useampaan tekstilajiin kuuluvien tekstien välillä on olemassa. Peruskysymyksiä ovat, millä tavoin merkityksiä välitetään tekstilajista toiseen, miten merkitykset muuttuvat tämän prosessin aikana ja mistä syystä. Metodisesti tällainen tutkimus edellyttää kahden tai useamman osa-aineiston keräämistä. Analyysivaiheessa näitä osa-aineistoja sitten vertaillaan toisiinsa ja osoitetaan, miten merkitykset muuttuvat niiden siirtyessä tekstilajista toiseen. Muutosten syiden selvittäminen saattaa edellyttää tekstianalyysin lisäksi tekstilajien käyttäjien haastattelemista tai heidän toimintansa havainnointia. Tekstilajiketjuja on tutkittu esimerkiksi sen kannalta, miten asiantuntijaraporteista laaditaan lehdistötiedotteita ja edelleen uutisjuttuja päivälehtiin; tutkimuksen perusteella merkityksenmuutokset ovat yleisiä etenkin lehdistötiedotteista tehdyissä uutisjutuissa, joissa esimerkiksi suoraviivaistetaan tiedotteissa kerrottuja kausaalisuhteita ja kärjistetään määrää koskevia väitteitä (ks. Solin 2001).

Rekontekstualisoinnin tutkimuksessa on puolestaan tarkasteltu esimerkiksi sitä, millaisin käytäntein toimituksellisista haastatteluista muokataan sitaatteja lehtijuttuun (Haapanen 2017). Haapasen mukaan käytänteet ovat moninaiset – sitaatin ja haastattelun välinen suhde ei useinkaan ole yksioikoinen. Esimerkiksi monologisoinniksi nimetyssä muokkauskäytänteessä toimittajan ja haastateltavan välinen vuoropuhelu pelkistetään lehtijutussa yhdeksi, haastateltavan tuottamaksi puheenvuoroksi. Tällöin haastateltavan äänellä saatetaan siis esittää myös sellaista, mitä hän ei tosiasiallisesti ole sanonut. (Mt.)

**Tekstianalyysin lähestymistapoja**

Tekstianalyysissa yhdistyy interaktionistinen ja konstruktivistinen tutkimusote: tekstejä tarkastellaan sekä paikallisena, tilannesidonnaisena vuorovaikutuksena että yhteisöllisenä ja kulttuurisena toimintana.

Tekstejä voidaan tutkia sekä tuotoksina (mm. kokonaisjäsentely) että prosesseina (mm. teeman- ja informaationkulku). Näkökulmat täydentävät toisiaan.

Kielen ja tekstin suhdetta voidaan tarkastella lähtemällä liikkeelle tietystä kielen keinosta ja tutkimalla sen käyttötapoja ja -yhteyksiä tekstijoukossa. Vaihtoehtoisesti voidaan ottaa lähtökohdaksi tietty toiminto tai merkityspiirre ja selvittää, kuinka se reaalistuu eli todentuu kielellisesti.

# 5. Yhteenveto

Nykyinen digitaalinen ja visuaalisia keinoja hyödyntävä tekstimaailma – niin sanottu uusi kirjoittaminen (engl. *new writing*; van Leeuwen 2008) – versoo moneen suuntaan, mikä vaatii tekstintutkimuksen menetelmien ja teorioiden jatkuvaa kehittämistä. Nykytekstit yhdistelevät yhä monipuolisemmin erilaisia semioottisia järjestelmiä ja moodeja toisiinsa ja hyödyntävät luovasti teknologisten alustojensa käyttömahdollisuuksia. Niiden koherenssi ei ehkä rakennu samalla tavoin kuin suunniteltujen monologimuotoon laadittujen tekstien. On myös olemassa digitaalisia kirjoitetun kielenkäytön muotoja, jotka asettuvat monologin ja vuoropuhelun välimaastoon. Esimerkiksi keskustelu- ja kommenttipalstojen vuorovaikutuksessa puheenvuorot ovat monesti rakenteellisesti ja toiminnallisesti mutkikkaampia kokonaisuuksia kuin pikaviestittelyn puheenvuorot, ja niissä käytetään monia perinteisille teksteille ominaisia jäsentelykeinoja, kuten otsikointia ja kappalejakoa. Tällaisten keskusteluiden tutkiminen teksti- ja keskustelunanalyysia yhdistelmällä voi olla perusteltu ratkaisu (ks. Virtanen & Kääntä 2018; Lahti 2019).

Sen lisäksi, että tekstintutkimus tuottaa tietoa teksteistä ja niiden moniulotteisista merkityksistä, samalla syntyy tietoa myös kielen osajärjestelmistä ja kieliopista. Esimerkiksi lauseyhdistelmien tutkiminen eri tekstilajeista on koetellut kieliopintutkimuksen näkemyksiä alisteisuudesta (ks. esim. Shore 2005; Juvonen 2010; Virtanen 2015, luku 4).

Monissa tutkimuksissa on kuvattu myös kulttuurisia kirjoittamisen käytänteitä ja yhteisöjen toimintaa yksittäisiä tekstejä tai genrejä laajemmasta näkökulmasta. Esimerkiksi oppikirjatekstien tutkimuksen avulla voidaan esittää huomioita pedagogisen tiedon luonteesta (esim. Karvonen 1995), samoin tutkimusartikkelien analyysi antaa kuvaa kulttuurieroista (esim. Mauranen 1993). Viranomaistekstejä tutkimalla voidaan kuvata hallinnon ideologiaa (esim. Sarangi & Slembrouck 1996; Tiililä 2007; Honkanen 2012).

Metateoreettisemmin suuntautunut tekstintutkimus on suunnannut kriittisen katseensa tutkimusalan ontologisiin, epistemologisiin ja metodologisiin perusoletuksiin (ks. Phakiti & Paltridge 2015, 17–20). Tästä hyviä esimerkkejä ovat pohdinnat siitä, mitä tekstit (ks. Visakko 2015) ja genret (ks. Nieminen 2010) ylipäätään ovat ja miten eri aikakausien aineistoja asetetaan samalle viivalle ottaen huomioon tekstien sosiokulttuuriset kontekstit (ks. Nord 2008; vrt. Näkökulmia filologiseen tutkimukseen tk.). Tekstintutkimus tarjoaa menetelmiä myös yhteiskunnan ja käyttäytymisen tutkijoille, joiden mielenkiinto koskee yhteiskunnan, vallan, demokratian, instituutioiden ja identiteettien tutkimusta eikä ensisijaisesti kielen ja kielenkäytön olemusta (ks. esim. Ruohonen, Sevänen & Turunen 2011; vrt. Diskurssianalyysi tk.). Näin tekstintutkimus elää omaa rikasta elämäänsä myös kielentutkimuksen ulkopuolella, mikä tarkoittaa jatkuvaa tutkimusalojen risteämistä.

**Aiheesta lisää:**

Bhatia, Vijay. 2004. *Worlds of Written Discourse. A Genre-based View*. London: Continuum.

Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä, Ulla & Lounela, Mikko (toim.). 2012. *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus.

Hiidenmaa, Pirjo. 2000. Lingvistinen tekstintutkimus. Julkaisussa: Sajavaara, Kari & Piirainen-Marsh, Arja (toim.) *Kieli, diskurssi ja yhteisö*. Jyväskylä: Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus, 161–190.

Holmberg, Per, Karlsson, Anna-Malin & Nord, Andreas (toim.). 2019 [2011]. *Funktionell textanalys*. Lund: Studentlitteratur.

Johnstone, Barbara. 2008 [2002]. *Discourse Analysis*. (2. uud. laitos). Malden (MA): Blackwell.

Martin, James R. & Rose, David. 2007 [2003]. *Working with Discourse*. (2. uud. laitos). London: Equinox.

Mäntynen, Anne, Shore, Susanna & Solin, Anna (toim.). 2006. *Genre – tekstilaji*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Pietikäinen, Sari & Mäntynen, Anne. 2019. *Uusi kurssi kohti diskurssia*. Tampere: Vastapaino.

## VIITTEET

1 Vuoropuheluna toteutuvien puhuttujen tekstien tutkimuksesta ks. Keskustelunanalyysi ja vuorovaikutuslingvistiikka tk.; vrt. Eggins & Slade (2006 [1996]). Kirjoitetun vuoropuhelun tutkimuksen menetelmistä ks. esim. Virtanen & Kääntä (2018). Internetin ei-lineaaristen hypertekstien tutkimuksesta ks. Koskimaa (2012).

2 Artikkelissamme esiteltyä tekstintutkimusta täydentää esimerkiksi tekstien kognitiivista prosessointia käsittelevä tutkimus, jossa tukeudutaan usein kokeellisiin asetelmiin (esim. Sanders, Schilperood & Spooren 2001; Kaakinen & Hyönä 2012).

3 *Diskurssilla* eli puhetavalla tarkoitetaan kiteytynyttä ja kulttuuris-yhteisöllisesti tunnistettavaa merkityksellistämisen tapaa, joka voi ilmetä esimerkiksi tieteellisille teksteille ominaisena näkökulmaisuutena (esim. Pietikäinen & Mäntynen 2019, 33; ks. myös Foucault 2005 [1969] ja Diskurssianalyysi tk.).

4 https://www.kielipankki.fi.

5 Tekstin pintatasosta voi tarkkailla jopa fonologisia ominaisuuksia, mikäli ne tuntuvat aineiston kannalta relevanteilta. Esimerkiksi runoudelle tyypillinen alkusointu on monen mainoksen tehokeino (”piristä päivääsi puristeilla”).

6 *Epistemologian* ja *ontologian* käsitteistä ks. Johdannon luku 2 tk.

7 Diskurssikuviot on alan kirjallisuudessa tapana kirjoittaa kapiteelein.

8 Lisäksi systeemis-funktionaalisessa perinteessä käytetään nimitystä *kytkentä* (engl. *conjunction*). Ks. Martin & Rose (2007); suomeksi Shore (2012).

9 Tekstilajin muutoksen tutkimuksesta ks. Kankaanpää (2012).

10 http://www.genreacrossborders.org.

## KIRJALLISUUS


Alasuutari, Pertti. 2011 [1993]. *Laadullinen tutkimus 2.0*. Tampere: Vastapaino.

Austin, John L. 2016 [1962]. *Näin tehdään sanoilla. Harvardissa 1955 pidetyt William James -luennot*. Alkuperäisteos *How to Do Things with Words* 1962. Suomentanut Risto Koskensilta. Tampere: niin & näin.

Bahtin, Mihail M. 1981 [1934–1935]. Discourse in the novel. Julkaisussa: Holquist, Michael (toim.) *The Dialogic Imagination. Four Essays*. Kääntäneet venäjästä englanniksi Caryl Emerson ja Michael Holquist. Austin: University of Texas Press, 259–422.

Bartlett, Tom. 2017. Positive discourse analysis. Julkaisussa: Flowerdew, John & Richardson, John (toim.) *The Routledge Handbook of Critical Discourse Studies*. London: Routledge, 133–147.

Bateman, John, Wildfeuer, Janina & Hiippala, Tuomo. 2017. *Multimodality. Foundations, Research and Analysis. A Problem-oriented Introduction*. Berlin: De Gruyter Mouton.

Beaugrande, Robert de & Dressler, Wolfgang. 1981. *Introduction to Text Linguistics*. London: Longman.

Berger, Peter & Luckmann, Thomas. 1994 [1966]. *Todellisuuden sosiaalinen rakentuminen: Tiedonsosiologinen tutkielma*. Suomentanut ja toimittanut Vesa Raiskila. Helsinki: Gaudeamus.

Bhatia, Vijay. 2004. *Worlds of Written Discourse. A Genre-based View*. London: Continuum.

Biber, Douglas. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Chafe, Wallace (toim.). 1980. *Pear Stories. Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Norwood (NJ): Ablex.

Devitt, Amy. 1991. Intertextuality in tax accounting: Generic, referential, and functional. Julkaisussa: Bazerman, Charles & Paradis, James (toim.) *Textual Dynamics of the Professions. Historical and Contemporary Studies of Writing in Professional Communities*. Madison (WI): University of Wisconsin Press, 336–357.

Dijk, Teun van. 1972. *Some Aspects of Text Grammars*. Hague: Mouton de Gruyter.

Eggins, Suzanne & Slade, Diana. 2006 [1996]. *Analysing Casual Conversation*. London: Equinox.

Enfield, Nick, Stivers, Tanya & Levinson, Stephen (toim.). 2010. Question-response sequences in conversation across ten languages. *Journal of Pragmatics* 42:10, 2615–2860.

Enkvist, Nils Erik. 1975. *Tekstilingvistiikan peruskäsitteitä*. Helsinki: Gaudeamus.

Fairclough, Norman. 2010 [1995]. *Critical discourse analysis*. (2. uud. laitos). Harlow: Pearson education.

Firth, John R. 1957 [1934–1950]. *Papers in Linguistics 1934–1951*. Oxford: Oxford University Press.

Foucault, Michel. 2005 [1969]. *Tiedon arkeologia*. Suom. Tapani Kilpeläinen. Tampere: Vastapaino.

Fries, Peter H. 1994. On theme, rheme and discourse goals. Julkaisussa: Coulthard, Malcolm (toim.) *Advances in Written Text Analysis*. London: Routledge, 229–249.

Haapanen, Lauri. 2017. *Quoting Practices in Written Journalism*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:ISBN 978-951-51-2952-9.

Hakulinen, Auli. 1982. Mikä tekee lauseista/virkkeistä tekstin. Julkaisussa: Tommola, J. & Ruusuvuori, R. (toim.) *Näkökulmia kirjoitettuun kieleen*. Jyväskylä: Suomen sovelletun kielitieteen yhdistys, 5–12.
Hakulinen, Auli & Karlsson, Fred. 1979. *Nykysuomen lauseoppia*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Hakulinen, Auli, Karlsson, Fred & Vilkuna, Maria. 1980. *Suomen tekstilauseiden piirteitä. Kvantitatiivinen tutkimus*. Helsinki: Helsingin yliopiston yleisen kielitieteen laitos.
Hakulinen, Auli & Selting, Margaret. 2005. Introduction. Julkaisussa: Hakulinen, Auli & Selting, Margaret (toim.) *Syntax and Lexis in Conversation. Studies on the Use of Linguistic Resources in Talk-in-interaction*. Amsterdam: John Benjamins, 1–14.
Halliday, Michael. 1978. *Language as Social Semiotic. The Social Interpretation of Language and Meaning*. London: Edward Arnold.
Halliday, Michael & Hasan, Ruqaiya. 1976. *Cohesion in English*. London: Longman.
Heikkilä, Elina. 2006. *Kuvan ja tekstin välissä. Kuvateksti uutiskuvan ja lehtijutun elementtinä*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Heikkinen, Vesa. 1999. *Ideologinen merkitys kriittisten tekstintutkimuksen teoriassa ja käytännössä*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Heikkinen, Vesa (toim.). 2009. *Kielen piirteet ja tekstilajit. Vaikuttavia valintoja tekstistä toiseen*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Heikkinen, Vesa, Hiidenmaa, Pirjo & Tiililä, Ulla. 2001. *Teksti työnä, virka kielenä*. Helsinki: Gaudeamus.
Heikkinen Vesa, Lounela, Mikko & Voutilainen, Eero. 2012. Aineistot ja niiden käyttö tekstilajitutkimuksessa. Julkaisussa: Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä, Ulla & Lounela, Mikko (toim.) *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus, 326–351.
Heikkinen, Vesa & Voutilainen, Eero. 2009. Koloratiivirakenne Hirventappopaikka-romaanissa. Julkaisussa: Heikkinen, Vesa (toim.) *Kielen piirteet ja tekstilajit. Vaikuttavia valintoja tekstistä toiseen*. Helsinki: Suomalaisen Kirjallisuuden Seura, 135–169.
Heikkinen, Vesa & Voutilainen, Eero. 2012. Genre – monitieteinen näkökulma. Julkaisussa: Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä, Ulla & Lounela, Mikko (toim.) *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus, 17–48.
Hiidenmaa, Pirjo. 2000. Lingvistinen tekstintutkimus. Julkaisussa: Sajavaara, Kari & Piirainen-Marsh, Arja (toim.) *Kieli, diskurssi ja yhteisö*. Jyväskylä: Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus, 161–190.
Hiidenmaa, Pirjo. 2001. Paperista tietoa. Julkaisussa: Charles, Mirjaliisa & Hiidenmaa, Pirjo (toim.) *Tietotyön yhteiskunta – kielen valtakunta*. AFinLAn vuosikirja 2001. Jyväskylä: Suomen soveltavan kielitieteen seura, 388–407.
Hoey, Michael. 1983. *On the Surface of Discourse*. London: George, Allen & Unwin.
Hoey, Michael. 2001. *Textual Interaction. An Introduction to Written Discourse Analysis*. London: Routledge.
Holmberg, Per, Karlsson, Anna-Malin & Nord, Andreas (toim.). 2019 [2011]. *Funktionell textanalys*. Lund: Studentlitteratur.

Honkanen, Suvi. 2009. Lauseet, jaksot ja (ala)laji viraston ohjailevissa kirjeissä. Julkaisussa: Heikkinen, Vesa (toim.) *Kielen piirteet ja tekstilajit. Vaikuttavia valintoja tekstistä toiseen*. Helsinki: Suomalaisen Kirjallisuuden Seura, 191–217.
Honkanen, Suvi. 2012. *Kielioppi ja tekstilaji. Direktiivin muotoilusta viraston ryhmäkirjeissä*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-7731-9.
Hyland, Ken. 2015. Researching writing. Julkaisussa: Paltridge, Brian & Phakiti, Aek (toim.) *Research Methods in Applied Linguistics*. London: Bloomsbury, 335–348.
Hymes, Dell. 1972. On communicative competence. Julkaisussa: Pride, John B. & Holmes, Janet (toim.) *Sociolinguistics*. Harmondsworth: Penguin, 277–284.
Johansson, Marjut, Nuolijärvi, Pirkko & Pyykkö, Riitta (toim.). 2011. *Kieli työssä. Asiantuntijatyön kielelliset käytännöt*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Johnstone, Barbara. 2008 [2002]. *Discourse Analysis*. (2. uud. p.). Malden (MA): Blackwell.
Juvonen, Riitta. 2010. Evaluoiva *että*-yhdyslause ja retoriset rakenteet suomenkielisessä ylioppilasaineessa. *Virittäjä* 114:1, 39–70.
Juvonen, Riitta. 2014. Näkökulma kirjoitelman dialogisuuteen. *Kuinka-* ja *miten-* yhdyslauseet ylioppilasaineessa. *Virittäjä* 118:1, 72–106.
Kaakinen, Johanna & Hyönä, Jukka. 2012. Kognitiivinen psykologia. Julkaisussa: Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä, Ulla & Lounela, Mikko (toim.) *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus, 546–550.
Kalliokoski, Jyrki. 1987. Kaunokirjallinen teos tekstien kosmoksessa. *Kirjallisuudentutkijain Seuran vuosikirja* 41, 179–198.
Kalliokoski, Jyrki. 1996a. Johdanto. Julkaisussa: Kalliokoski, Jyrki (toim.) *Teksti ja ideologia. Kirjoituksia kielestä ja vallasta julkisessa kielenkäytössä*. Kieli 9. Helsinki: Helsingin yliopiston suomen kielen laitos, 8–36.
Kalliokoski, Jyrki (toim.). 1996b. *Teksti ja ideologia. Kirjoituksia kielestä ja vallasta julkisessa kielenkäytössä*. Kieli 9. Helsinki: Helsingin yliopiston suomen kielen laitos.
Kalliokoski, Jyrki. 2008. Kirjailijan kieltä tutkimassa. Julkaisussa: Siiroinen, Mari & Onikki-Rantajääskö, Tiina (toim.) *Kieltä kohti*. Helsinki: Otava, 34–48.
Kankaanpää, Salli. 2006. *Hallinnon lehdistötiedotteiden kieli*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Kankaanpää, Salli. 2012. Tekstilajin muutoksen laadullinen tutkiminen. Julkaisussa: Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä, Ulla & Lounela, Mikko (toim.) *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus, 285–296.
Karvonen, Pirjo. 1995. *Oppikirjateksti toimintana*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Kauppinen, Anneli & Laurinen, Leena. 1984. *Tekstejä teksteistä. Muisti- ja tekstilingvistiikan sovellutuksia asiatekstien referoinnin problematiikkaan*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Koivusalo, Esko & Huovinen-Nyberg, Liisa. 1978. *Selkeä virkakieli*. Helsinki: Valtion painatuskeskus.
Komppa, Johanna. 2012. *Retorisen rakenteen teoria suomi toisena kielenä -ylioppilaskokeen kirjoitelman kokonaisrakenteen ja kappalejaon tarkastelussa*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-8164-4.
Koskela, Sanni. 2014. Työhakemus tekstilajina. *Virittäjä* 121:1, 5–40.

Koskimaa, Raine. 2012. Hypertekstintutkimus. Julkaisussa: Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä, Ulla & Lounela, Mikko (toim.) *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus, 511–515.
Kress, Günther. 2009. *Multimodality. A social semiotic approach to contemporary communication*. London: Routledge.
Kristeva, Julia. 1993 [1967]. Sana, dialogi ja romaani. Julkaisussa: Kristeva, Julia, *Puhuva subjekti. Tekstejä 1967–1993*. Suom. Pia Sivenius, Tiina Arppe, Kirsi Saarikangas, Helena Sinervo ja Riikka Stewen. Helsinki: Gaudeamus, 21–50.
Kuiri, Kaija. 1984. *Referointi Kainuun ja Pohjois-Karjalan murteissa*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Kuortti, Joel, Mäntynen, Anne & Pietikäinen, Sari. 2008. Kielen rakennustelineillä. Kielellisen ja yhteiskunnallisen käänteen merkitys. *Tiedotustutkimus* 31:3, 25–37.
Labov, William & Waletzky, Joshua. 1967. Narrative analysis. Oral versions of personal experience. Julkaisussa: Helm, June (toim.) *Essays on the Verbal and Visual Arts*. Washington (ST): University of Washington Press, 12–44.
Lahti, Emmi. 2019. *Maahanmuuttokeskustelun retoriikkaa*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-5707-2.
Leeuwen, Theo van. 2008. New forms of writing, new visual competencies. *Visual Studies* 23:2, 130–135.
Lehtinen, Esa & Pälli, Pekka. 2016. Osallistuminen näkökulmana tekstien ja keskustelujen suhteeseen: Kehityskeskustelulomake osana kehityskeskustelun vuorovaikutusta ja organisatorisia käytänteitä. *Virittäjä* 120:2, 164–187.
Lillis, Theresa. 2013. *Sociolinguistics of Writing*. Edinburgh: Edinburgh University Press.
Makkonen-Craig, Henna. 2005. *Toimittajan läsnäolo sanomalehtitekstissä. Näkökulmia suomen kielen dialogisiin passiivilauseisiin*. SKST 1026. Helsinki: Suomalaisen Kirjallisuuden Seura.
Mann, William & Thompson, Sandra. 1988. Rhetorical structure theory. Toward a functional theory of text organization. *Text* 8, 243–281.
Martin, James R. 1992. *English text. System and Structure*. Amsterdam: John Benjamins.
Martin, James R. & Rose, David. 2007 [2003]. *Working with Discourse*. (2. uud. laitos). London: Equinox.
Martin, James R. & Rose, David. 2008. *Genre Relations. Mapping Culture*. London: Equinox.
Martin, James R. & White, Peter R. R. 2005. *The Language of Evaluation. Appraisal in English*. London: Palgrave Macmillan.
Mauranen, Anna. 1993. *Cultural Differences in Academic Rhetoric. A Textlinguistic Study*. Frankfurt am Main: Peter Lang.
Mikkola, Piia. 2016. *Teksti, agenda, artefakti. Kehityskeskustelulomakkeen erilaiset roolit ja kehityskeskustelujen topikaalisissa siirtymissä*. Väitöskirja. Vaasa: Vaasan yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-476-689-0:.
Miller, Carolyn. 1984. Genre as social action. *Quarterly Journal of Speech*, 151–167.
Miller, Donna R. & Bayley, Paul (toim.). 2016. *Hybridity in Systemic Functional Linguistics. Grammar, Text and Discursive Context*. London: Equinox.
Mäntynen, Anne. 2003. *Miten kielestä kerrotaan. Kielijuttujen retoriikkaa*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Mäntynen, Anne. 2005. Referointi tekstilajipiirteenä. Esimerkkinä kielijutut. Julkaisussa: Haakana, Markku & Kalliokoski, Jyrki (toim.) *Referointi ja moniäänisyys*. Tietolipas 206. Helsinki: Suomalaisen Kirjallisuuden Seura, 258–281.
Mäntynen, Anne. 2006. Näkökulmia tekstin ja tekstilajien rakenteeseen. Julkaisussa: Mäntynen, Anne, Shore, Susanna & Solin, Anna (toim.) *Genre – tekstilaji*. Helsinki: Suomalaisen Kirjallisuuden Seura, 42–71.
Mäntynen, Anne & Shore, Susanna. 2014. What is meant by hybridity? An investigation of hybridity and related terms in genre studies. *Text & Talk* 6, 737–758.
Nieminen, Tommi. 2010. *Lajien synty*. Väitöskirja. Jyväskylä: Jyväskylän yliopisto.
Nord, Andreas. 2008. *Trädgårdsboken som text 1643–2005*. Väitöskirja. Tukholma: Tukholman yliopisto.
Phakiti, Aek & Paltridge, Brian. 2015. Approaches and methods in applied linguistic research. Julkaisussa: Paltridge, Brian & Phakiti, Aek (toim.) *Research Methods in Applied Linguistics*. London: Bloomsbury, 5–25.
Pietikäinen, Sari & Mäntynen, Anne. 2019. *Uusi kurssi kohti diskurssia*. Tampere: Vastapaino.
Raevaara, Liisa. 1997. Vierusparit – esimerkkinä kysymys ja vastaus. Julkaisussa: Tainio, Liisa (toim.) *Keskustelunanalyysin perusteet*. Tampere: Vastapaino, 75–92.
Rahtu, Toini. 2006. *Sekä että. Ironia koherenssina ja inkoherenssina*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Rahtu, Toini. 2012. Lingvistinen reseptioanalyysi tekstilajitutkijan palveluksessa. Julkaisussa: Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä, Ulla & Lounela, Mikko (toim.) *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus, 433–440.
Rorty, Richard. 1967. *The Linguistic Turn. Recent Essays in Philosophical Method*. Chicago (IL): University of Chicago Press.
Ruohonen, Voitto, Sevänen, Erkki & Turunen, Risto (toim.). 2011. *Paluu maailmaan. Kirjallisten tekstien sosiologiaa*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Sanders, Ted, Schilperood, Joost & Spooren, Wilbert (toim.). 2001. *Text Representation. Linguistic and Psycholinguistic Aspects*. Amsterdam: John Benjamins.
Sarangi, Srikant & Slembrouck, Jeff. 1996. *Language, Bureaucracy and Social Control*. London: Longman.
Salomaa, Elina. 2019. *Television ja Twitterin risteyksessä: Sosiaalinen televisio vuorovaikutuksen ja mediatapahtumaan osallistumisen välineenä*. Väitöskirja. Jyväskylä: Jyväskylän yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-39-7869-3.
Saukkonen, Pauli. 1984. *Mistä tyyli syntyy?* Helsinki: WSOY.
Saussure, Ferdinand de. 2014 [1916]. *Yleisen kielitieteen kurssi* [*Cours de linguistique générale*]. Suom. Tommi Nuopponen. Tampere: Vastapaino.
Searle, John. 1969. *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
Shore, Susanna. 2005. Referoinnista projektioon ja metarepresentaatioon. Julkaisussa: Kalliokoski, Jyrki & Haakana, Markku (toim.) *Referointi ja moniäänisyys*. Helsinki: Suomalaisen Kirjallisuuden Seura, 44–82.
Shore, Susanna. 2008. Lauseiden tekstuaalisesta jäsennyksestä. *Virittäjä* 112:1, 24–65.
Shore, Susanna. 2012. Kieli, kielenkäyttö ja kielenkäytön lajit systeemis-funktionaalisessa teoriassa. Julkaisussa: Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä,

Ulla & Lounela, Mikko (toim.) *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus, 131–157.
Shore, Susanna & Mäntynen, Anne. 2006. Johdanto. Julkaisussa: Mäntynen, Anne, Shore, Susanna & Solin, Anna (toim.) *Genre – tekstilaji*. Helsinki: Suomalaisen Kirjallisuuden Seura, 9–41.
Solin, Anna. 2001. *Tracing Texts. Intertextuality in Environmental Discourse*. Väitöskirja. Helsinki: Helsingin yliopisto.
Solin, Anna. 2009. Genre. Julkaisussa: Östman, Jan-Ola Östman & Verschueren, Jef (toim.) *Handbook of Pragmatics*. Amsterdam: John Benjamins.
Swales, John. 1990. *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.
Tainio, Liisa & Karvonen, Ulla. 2015. Finnish teachers exploring gender bias in school textbooks. Julkaisussa: Mills, Sara & Mustapha, Abolaji S. (toim.) *Gender Representations in Learning Materials. International Perspectives*. London: Routledge, 124–149.
Thompson, Geoff. 2001. Interaction in academic writing. Learning to argue with the reader. *Applied Linguistics* 22, 58–78.
Thompson, Geoff. 2007. Unfolding theme. The development of clausal and textual perspectives on theme. Julkaisussa: Webster, Jonathan, Hasan, Ruqaiya & Matthiessen, Christian (toim.) *Continuing Discourse on Language. Volume 2: A Functional Perspective*. London: Equinox, 671–696.
Tiililä, Ulla. 2007. *Tekstit viraston työssä. Tutkimus etuuspäätösten kielestä ja konteksteista*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Tiililä, Ulla. 2014. Verbaaliset sormenjäljet – kielentutkimus huijausten ja rikosten tutkinnassa. *Kielikello* 4/2014, 14–17.
Tiililä, Ulla & Karvinen, Kati (toim.). 2017. *Elämän ja kuoleman tekstit. Kirjoittaminen sosiaali- ja hoitotyössä*. Helsinki: Kotimaisten kielten keskus. [verkkoaineisto]. [viitattu 15.4.2018]. Saatavissa: http://scripta.kotus.fi/www/verkkojulkaisut/julk50/.
Tiittula, Liisa. 2012. Haastattelut tekstilajitutkimuksessa. Julkaisussa: Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä, Ulla & Lounela, Mikko (toim.) *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus, 423–432.
Uotinen, Suvi. 2007. ”Tässä kohdassa heterorouva nyökkää.” Näkökulma naisparien henkilöhaastatteluissa. *Virittäjä* 111:1, 68–88.
Virtanen, Mikko T. 2015. *Akateeminen kirja-arvio moniäänisenä toimintana*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-1760-1.
Virtanen, Mikko T. & Juvonen, Riitta. 2018. Moniäänisyys, sekventiaalisuus ja kerronta. Alisteisen interrogatiivilauseen sisältävät rakenteet blogivuorovaikutuksessa. Julkaisussa: Rahtu, Toini, Shore, Susanna & Virtanen, Mikko T. (toim.) *Kirjoitettu vuorovaikutus*. Helsinki: Suomalaisen Kirjallisuuden Seura, 179–213.
Virtanen, Mikko T. & Kääntä, Liisa. 2018. At the intersection of text and conversation analysis. Analysing asynchronous online written interaction. *AFinLA-e: Soveltavan kielitieteen tutkimuksia* 10, 140–158.
Virtanen, Mikko, Rahtu, Toini & Shore, Susanna. 2018. Kirjoitetun vuorovaikutuksen moninaisuus ja erityisyys. Julkaisussa: Rahtu, Toini, Shore, Susanna & Virtanen, Mikko T. (toim.) *Kirjoitettu vuorovaikutus*. Helsinki: Suomalaisen Kirjallisuuden Seura, 9–38.

Visakko, Tomi. 2015. *Self-promotion as Semiotic Behaviour. The Mediation of Personhood in Light of Finnish Dating Advertisements*. Väitöskirja. Helsinki: Helsingin yliopisto.
Visakko, Tomi. 2018. Päämäärät, ongelmat ja tilaisuudet. Eksistentiaalinen rationaalisuus ja kausaalis-evaluatiiviset diskurssikuviot yritysten verkkosivujen visio-, missio- ja strategiateksteissä. *Sananjalka* 60, 142–161.
Visakko, Tomi & Voutilainen, Eero. 2012. Aprillipila tekstilajina. Julkaisussa: Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä, Ulla & Lounela, Mikko (toim.) *Genreanalyysi – tekstilajitutkimuksen käytäntöä*. Helsinki: Kotimaisten kielten keskus, 265–301.
Vitikka, Elina. 2018. Hyperlinkkien funktiot ja kirjoittaja- ja lukijapositiot. *Virittäjä* 122, 161–186.
Vuorijärvi, Aino. 2013. *Tekstilaji ja yhteisö. Ammattikorkeakoulun opinnäytetyön diskussio tekstinä*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-8928-2.
Werlich, Egon. 1975. *Typologie der Texte. Entwurf eines textlinguistichen Modells zur Grundlegung einer Textgrammatik*. Heidelberg: Quelle & Meyer.
Winter, Eugene. 1977. A clause-relational approach to English texts: A study of some predictive lexical items in written discourse. *Instructional Science* 6:1, 1–92.
Winter, Eugene. 1994. Clause relations as information structure: Two basic text structures in English. Julkaisussa: Coulthard, Malcolm (toim.) *Advances in Written Text Analysis*. London: Routledge, 46–68.

# Luku 21 Sosiolingvistiikka

*Pirkko Nuolijärvi*
https://orcid.org/0000-0002-9688-6314
*Hanna Lappalainen*
https://orcid.org/0000-0001-5422-1667

**Mikä?**

Sosiolingvistiikassa tutkitaan kielen ja yhteiskunnan suhdetta. Keskeisiä tutkimuskohteita ovat kielen variaatio ja muutos, variaation ja vuorovaikutuksen suhde, monikielisyys, kielenvaihto ja kielen säilyttäminen sekä kieliasenteet. Tutkimuksessa sovelletaan sekä kvantitatiivisia että kvalitatiivisia menetelmiä ja aineistot voivat olla monentyyppisiä.

**Katso myös:**

Luku 2 Kysely

Luku 3 Haastattelu aineistonkeruun metodina sosiolingvistiikassa

Luku 5 Lingvistinen etnografia toimintayhteisöissä

Luku 22 Keskustelunanalyysi ja vuorovaikutuslingvistiikka

# 1. Johdanto: mitä on sosiolingvistiikka?

Sosiolingvistiikka on kielentutkimuksen alue, joka tarkastelee kielenkäytön ja yhteiskunnan sosiaalisen rakentumisen suhdetta. Tutkimuskohteina ovat kieliyhteisön variaatio, kielen muuttuminen, kielen funktiot vuorovaikutuksessa, kieliä koskevat asenteet ja käsitykset sekä monikielisyys. Historiallinen sosiolingvistiikka tarkastelee kielenmuutosta pitemmällä aikavälillä ja kirjallisten lähteiden avulla (ks. esim. Nevalainen & Raumolin-Brunberg 2003), kun taas nykykieleen keskittyvä sosiolingvistinen tutkimus lähestyy muutosta ennen muuta synkronisesti tai ainakin lyhyemmällä aikajänteellä puhutun ja kirjoitetun kielen aineistojen valossa. Diakronisen ja synkronisen näkökulman raja ei kuitenkaan ole jyrkkä, koska kielenkäytön tutkiminen ja muutosten tarkastelu tiettynä ajankohtana tarvitsee taustakseen tietoa aiempien vaiheiden kielenkäytöstä (ks. esim. Milroy, J. & Milroy, L. 1997).

Sosiolingvistiikka on siten vahvasti aineistolähtöinen tutkimussuunta, joka määrällisin ja laadullisin menetelmin analysoi kielellisen variaation esiintymisehtoja, variaation funktioita sekä niiden muutosta yhteiskunnallisessa kontekstissa. Analyysissa on olennaista osoittaa, mihin aineistosta tehdyt tulkinnat perustuvat. Sosiolingvistiikka on siis empiiriseen näyttöön perustuvaa tutkimusta, jossa aineiston rajallisuus tai käsittelytavan kriittiset pisteet tuodaan lukijan näkyville. (Haastatteluaineiston keruun, käsittelyn ja julkaisemisen etiikasta ks. Haastattelu aineistonkeruun metodina sosiolingvistiikassa tässä kirjassa.)

Sosiolingvistiikkaa ja kielisosiologiaa on usein pidetty synonyymeina, mutta niitä edustavat suuntaukset painottuvat eri tavoin. Teoksessaan *Sociolinguistic Theory* (2009) J. K. Chambers määrittelee sosiolingvistiikan ja kielisosiologian limittyvien kenttien eron niin, että kummassakin suuntauksessa tutkitaan kielen ja yhteiskunnan suhdetta, mutta sosiolingvistiikan tavoitteena on ymmärtää kielen rakennetta ja kielisosiologian tavoitteena yhteiskunnan rakennetta. Sosiolingvistiikan myöhempi historia on osoittanut, että tämä ero ei ole selvärajainen; yksilöiden ja ryhmien kielen vaihtelua tutkiessaan sosiolingvisti on mitä suurimmassa määrin yhteiskunnan rakenteen ytimessä. Chambersin

kuvaus sosiolingvistiikan alasta on kaikkineenkin suppea. Laajimmin sosiolingvistiikan alaa lienee kuvannut M. A. K. Halliday, joka jo vuonna 1972 esitti, että sosiolingvistiikkaan voi lukea seuraavat alueet (ks. Gardin & Marcellesi 1987): kielen makrososiologia ja kielellinen demografia; diglossia, monikielisyys ja useamman kielimuodon hallinta; kielen suunnittelu, kehitys ja standardointi; pidginisaatio ja kreolisaatio; sosiaalisten murteiden ja ei-standardien varieteettien kuvaus; sosiolingvistiikka ja koulutus; puheen etnografia; rekisterit ja repertuaarit sekä koodinvaihto; kieliopillisten ja fonologisten muutosten sosiaaliset taustatekijät; kieli, sosiaalistuminen ja kulttuurinen muutos; sosiolingvistinen näkökulma lasten kielelliseen kehitykseen; kielellisen systeemin funktionaaliset teoriat; lingvistinen relativismi; etnometodologinen lingvistiikka ja tekstiteoria. Nykynäkökulmasta sosiolingvistinen lähestymistapa koskettaa monia näistä aloista, vaikkei kaikkia yhtä vahvasti.

Tässä artikkelissa keskitymme erityisesti niihin alueisiin, jotka ovat lähes 50 vuoden aikana olleet suomalaisessa tutkimuksessa keskeisiä. Vaikka monet sosiolingvistit niin Suomessa kuin muuallakin ovat olleet aktiivisia myös kielipolitiikan ja kielikoulutuspolitiikan alueella ja osallistuneet siihen liittyvään lainsäädäntötyöhön ja kielten asemaa koskevaan keskusteluun, usein yhdessä muiden alojen lingvistien kanssa (ks. esim. Latomaa & Nuolijärvi 2005; Gregersen 2014; Sarhimaa 2017), rajaamme tässä yhteydessä tämän kentän tarkastelumme ulkopuolelle.

## 2. Suomalaisen sosiolingvistiikan kansainvälistä taustaa ja historiaa

Ensimmäisen kerran *sosiolingvistiikka*-termi on tiettävästi mainittu vuonna 1952, jolloin Haver C. Currie julkaisi artikkelin ”Projections of sociolinguistics” (ks. Coulmas 1997). Sosiaaliset erot huomioon ottava tutkimus alkoi 1950-luvulla tulla osaksi kielentutkimusta, mutta käsitteenä ja tutkimussuuntana sosiolingvistiikka vakiintui vasta 1960-luvulla William Labovin (1966) työn ansiosta. Labovilainen tutkimustraditio

edustaa variationistista sosiolingvistiikkaa. Muita keskeisiä tutkimusperinteitä ovat antropologinen traditio (ks. Hymes 1971), vuorovaikutussosiolingvistiikka (ks. Gumperz 1982), kielisosiologia (ks. Fishman 1968) ja kieliasennetutkimus eri muodoissaan (ks. esim. Giles & St Clair 1979; Preston 1989; 1999; Long & Preston 2000).

Sosiolingvistiikan keskeisiä käsitteitä ovat *vaihtelu* ja *muutos*, ja sen piirissä näitä tarkastellaan yhteiskunnallisten ja sosiaalisten muutosten kehyksissä. Nikolas Coupland (2009) suhteuttaa sosiolingvistisen tutkimuksen taustan ja myöhemmät vaiheet yhteiskunnalliseen kehitykseen seuraavalla tavalla: Dialektologista tutkimusta harjoitettiin esimodernistisessa vaiheessa. Tällöin perinteinen dialektologia oli läheisesti sidoksissa historialliseen kielentutkimukseen ja kielellisten muotojen arveltiin todistavan vanhemmista kulttuurikäytänteistä. Sosiaaliset prosessit sinällään eivät kuitenkaan olleet murteentutkijoiden kiinnostuksen kohteina. Variationistinen sosiolingvistiikka puolestaan perustui modernismin käsityksiin yhteiskunnasta; kielenkäytön taustamuuttujina tarkasteltiin sukupuolta, luokkaa ja ikää. Ne vakiintuivat osaksi variaationtutkimusta, ja ne nähtiin ikään kuin annettuina (taustamuuttujia koskevasta kritiikistä ks. esim. Nuolijärvi 2000). Variationistinen sosiolingvistiikkakaan ei juuri tarkastellut sitä, miten kielen variaatio ja käyttö ovat sidoksissa muuttuvaan sosiokulttuuriseen ekosysteemiin. Nykyistä niin sanottua dekonstruktiivista sosiolingvistiikkaa harjoitetaan myöhäismodernismin ja globalisaation ympäristössä; *murteen* ja *standardin* käsitteitä tarkastellaan kriittisemmin kuin aiemmin ja selvärajaisten varieteettien olemassaolo kyseenalaistetaan (Johnstone 2011; Vaattovaara 2016, 85–92). Couplandin mukaan tähän mennessä sosiolingvistiikassa on kuitenkin kaikkina aikoina tarkasteltu kielen ja sosiaalisen muutoksen suhdetta varsin rajallisesti.

Sosiolingvistinen tutkimus on silti implisiittisesti aina lähtenyt siitä, että yhteiskunnan muutokset, kuten elinkeinorakenteen ja koulutuksen muutokset, näkyvät kielessä. Ne vaikuttavat kielenkäyttötapoihin ja vuorovaikutuksen kehyksiin, mutta tätä prosessia ei pidä olettaa yksisuuntaiseksi eikä mekaaniseksi. Kielenkäyttö puolestaan vaikuttaa siihen, millainen yhteiskunta on ja millaisena sen toiminta näyttäytyy, eikä näitä kahta voi pitää toisistaan erillisinä ilmiöinä.

Vaihtelu kuuluu kielen olemukseen, ja sen pohjalta kieli voi pitkällä aikavälillä muuttua, mutta kaikki variaatio ei johda muutokseen. Siten sosiolingvistiikka on kiinnostunut myös siitä, millainen vaihtelu on pysyvää, millaiset muodot elävät rinnakkain kielenkäyttäjien repertuaarissa eli yksilön tai ryhmän käyttämien kielimuotojen muodostamassa kokonaisuudessa ja mikä kunkin muodon tehtävä on vuorovaikutuksessa. Sosiolingvistiikka pyrkii selvittämään myös sitä, miksi jokin variantti jää pois käytöstä. Analyysissa kiinnitetään huomio sekä kielensisäisiin että -ulkoisiin tekijöihin. Kaikkiaan voi yhtyä Couplandin edellä kuvattuihin käsityksiin siitä, että sosiolingvistit ovat tulkinneet yhteiskunnan ja kielen suhdetta varsin varovaisesti, vaikka eri tutkijat ovat myös Suomessa painottaneet tätä suhdetta eri tavoin.

Suomalaisen sosiolingvistiikan alkuvaiheet nousevat sekä dialektologian tutkimuksesta että labovilaisen variaationtutkimuksen perinteestä (ks. esim. Siitonen 1981; Mielikäinen 1982; Juusela 1994). Dialektologit tutkivat alueellista vaihtelua analysoimalla lähinnä maaseudun murteiden äänne- ja muotopiirteitä. Kohteena oli samalla maaseutupaikkakunnalla asuneiden vanhusten kielimuoto. Labovin jalanjäljissä brittiläiset ja pohjoismaiset sosiolingvistit kohdistivat huomionsa kaupunkipuhekielten ja maaseudun murteiden fonetiikkaan, fonologiaan ja morfologiaan. Tutkimusten kohteena oli eri-ikäisten ja eritaustaisten puhujien kielimuoto. Ennen varsinaisen sosiolingvistisen tutkimuksen käynnistämistä Suomessa julkaistiin murteentutkimuksen piirissä muutamia tutkimuksia siitä, miten muutto vaikuttaa murteen muuttumiseen. Niissä analysoitiin eri paikkakunnille sodan jälkeen asutettujen eri-ikäisten siirtokarjalaisten puhekieltä (ks. Jahnukainen 1954; Leskinen 1974; 1980). Jahnukaisen pro gradu -tutkielmaa ja Leskisen varhaisempaa tutkimusta voi pitää eräänlaisina suomalaisen sosiolingvistisen tutkimuksen etiäisinä esimerkiksi siksi, että niissä tarkastellaan useamman sukupolven ja sukupuolten välisiä eroja, mikä sittemmin on ollut tavallista sosiolingvistisissä variaationtutkimuksissa.

Sosiolingvisti Penelope Eckert (2012) on hahmotellut kolme sosiolingvistiikan aaltoa, jotka ovat tunnistettavissa myös suomen kielen sosiolingvistisessä tutkimuksessa. Aallot ovat kuitenkin vain suuntaa antavia jäsennyksiä, eivätkä eri aaltojen rajat siten ole selviä; toisistaan

erottuvien aaltojen sijasta voi mieluummin nähdä aallokon (ks. esim. Eckert & Labov 2017). Numerointi synnyttää helposti ajatuksen aaltojen erilaisesta arvottamisesta, mutta oma lähtökohtamme on, että numerointi kuvaa aaltojen ajallista järjestystä eli osittain tutkimushistoriallista kronologiaa. Jokaisella vaiheella on oma olennainen paikkansa, eikä uuden vaiheen alku tarkoita edellisen loppumista. Kysymys on enemmänkin tutkimuksen painopisteiden muutoksesta tai uusien tutkimustapojen tulosta entisten rinnalle.

Ensimmäisessä aallossa vallitsevana suuntauksena oli kvantitatiiviseen analyysiin painottunut labovilainen sosiolingvistiikka, jonka mukaan ihminen omaksuu puhetapansa ympäristön vaikutuksesta (Labov 1966). Labovin lisäksi suomalaiseen tutkimukseen vaikutti vahvasti brittiläinen variaationtutkimus (Trudgill 1974) sekä pohjoismainen tutkimus ja sittemmin yhteistyö pohjoismaisten tutkijoiden kanssa (ks. esim. Thelander 1979a; 1979b; Nordberg 1994). Suomessa vallalla oli 1970-luvulta 1990-luvulle asti juuri labovilainen sosiolingvistiikka. Labovilaisen variaationtutkimuksen menetelmiä käytettiin Suomessa ensimmäisen kerran Nykysuomalaisen puhekielen murroksen tutkimuksessa (1976), jonka vastuullisena johtajana toimi Heikki Paunonen. Variaatiota suhteutettiin kielenulkoisiin muuttujiin, joista yleisimmin tarkasteltiin ikää, sukupuolta ja sosiaaliryhmää sekä asuinympäristöä. Painettujen julkaisujen lisäksi (ks. esim. Mielikäinen 1980; 1981b; 1981c; 1986; Jonninen-Niilekselä 1982; Suojanen, M. K. 1981; 1985; Nuolijärvi 1986b) runsaasti tutkimustietoa on kymmenissä pro gradu -tutkielmissa, joissa on pääosin tarkasteltu äänne- ja muotovariaatiota. Erityisen kiinnostuksen kohteena oli ikämuuttuja ja nuorten kieli, koska yhtenä keskeisenä oletuksena oli, että nuorten valinnat kuvastavat kielen muutoksen trendejä (ks. Lappalainen 2001). Samaan aikaan oli kuitenkin merkkejä myös toisenlaisista lähestymistavoista, kuten Turun puhekielen projektissa, jossa aineistona oli myös kokouskeskusteluja ja tutkimuksen kohteena syntaksin ilmiöitä (ks. mm. Suojanen, M. K. & Suojanen, P. 1982; Suojanen, M. K. 1985). Myös perinteinen kansanlingvistiikka voidaan lukea ensimmäisen aallon tutkimuksiin; se on tarkastellut kielellisiä asenteita ja käsityksiä kielimuodoista (ks. esim. Mielikäinen & Palander 2014; Vaattovaara 2009; Nupponen 2011).

Toisen aallon sosiolingvistiikassa korostuvat sosiaaliset verkostot, ryhmäkytkökset sekä sosiaalinen orientaatio ja kielellinen identiteetti (esim. Milroy 1980; Eckert 2000). Yksilöillä on erilaisia tyylillisiä käytänteitä, ja sosiaaliryhmällä ja sukupuolella on merkittävä rooli. Suomessa variaationanalyysiin on yhdistetty etnometodologisesti suuntautunutta vuorovaikutuksen tutkimusta (esim. Lappalainen 2004; ks. myös Keskustelunanalyysi ja vuorovaikutuslingvistiikka tk.).

Kolmannelle aallolle on tyypillistä, että variaatiota tarkastellaan indeksisenä järjestelmänä; kiinnostuksen kohteena ovat indeksiset kentät ja se, miten merkitys syntyy vuorovaikutustilanteessa. Taustalla on Michael Silversteinin (2003) indeksisen järjestyksen (engl. *indexical order*) pohjalta kehitetty ajatus sosiaalisesta indeksisyydestä, jolla tarkoitetaan sitä, että esimerkiksi jokin kielenpiirre synnyttää mielleyhtymän johonkin osallistujien tunnistamaan sosiaaliseen ryhmään niin, että tuosta assosiaatiosta tulee osa kyseisen kielenpiirteen tulkintaa (ks. esim. Lehtonen 2015, 35–46). Tutkimus on usein etnografisesti suuntautunutta (ks. tarkemmin Lingvistinen etnografia toimintayhteisöissä tk.). Siinä tarkastellaan, miten kielen avulla jäsennetään ja tuotetaan sosiaalista todellisuutta ja kiinnitetään huomiota kielenkäytön historiallisiin tai kulttuurisiin taustoihin sekä kieli-ideologioihin (esim. Lehtonen 2015; Sorjonen, Rouhikoski & Lehtonen 2015; Vaattovaara & Poutiainen 2017).

Sosiolingvistit ovat vuosikymmenten kuluessa tarkastelleet kieliyhteisöjä monenlaisissa konteksteissa erilaisten aineistojen valossa. Teoreettista keskustelua on käyty erityisesti kielen ja yhteiskunnan suhteen painottumisesta ja eri aikoina etenevien sosiaalisten prosessien läsnäolosta kielenkäytön muutosten analyysissa. Alan tutkimusta on esitelty lukuisissa teoksissa (ks. esim. Coupland 2007; 2016; Coupland & Jaworski 1997; Bayley, Cameron & Lucas 2013).

# 3. Sosiolingvististä tutkimusta 1970-luvulta 2010-luvulle

Seuraavissa luvuissa tarkastelemme suomalaisen sosiolingvistisen tutkimuksen eri vaiheita, lähestymistapoja ja menetelmiä – niiden kansainvälisiä esikuvia unohtamatta. Tarkastelu etenee osittain kronologisesti mutta tuo esiin myös sen, miten eri lähestymistavat ja menetelmät ovat kehittyneet ja miten samoja menetelmiä voi hyödyntää uusissa oloissa ja uusien aineistojen avulla – myös uusiin menetelmiin yhdistellen. Samoin aiemmin koottuja puhutun kielen aineistoja voi käsitellä uusista näkökulmista ja uusin menetelmin ja siten hyödyntää niitä uudelleen. Tarkastelumme punaisena lankana onkin se, että niin eri tilanteissa tallennettujen yksilö- ja ryhmähaastattelujen sekä muunlaisten vuorovaikutustilanteiden analysointi kuin kielenkäytön systemaattinen havainnointi tarjoavat sosiolingvistiselle tutkimukselle mahdollisuuden erilaisiin näkökulmiin ja painotuksiin. Lähdemme liikkeelle variaationtutkimuksesta ja etenemme kielenvaihdon ja kielensäilyttämisen analyysiin. Sen jälkeen käsittelemme vuorovaikutussosiolingvistiikan näkökulmia, kieliasenteiden tutkimusta ja monikielisten toimintayhteisöjen tutkimusta.

## 3.1. Variaatio ja kielenmuutos

Kaupunkipuhekielten variationistista tutkimusta on harjoitettu laajamittaisemmin 1960-luvun lopulta lähtien näihin päiviin asti erityisesti anglosaksisissa maissa (Labov 1966; Trudgill 1974), saksalaisella kielialueella (Kallmeyer 1987; 1994) ja pohjoismaissa (Nordberg 1971; 1994; Thelander 1979a; 1979b; Sundgren 2002). Aineistoa on koottu yleensä haastattelemalla useamman, tavallisesti kolmen, sukupolven kielenkäyttäjiä. Suomessa ensimmäinen laaja projekti oli edellä mainittu Nykysuomalaisen puhekielen murros, joka toteutettiin neljässä kaupungissa, Helsingissä, Tampereella, Turussa ja Jyväskylässä. Kaikissa ympäristöissä tutkittiin ensisijaisesti äänne- ja muotovariaatiota

ja tarkasteltiin koulutukseltaan eritaustaisia ja eri ammateissa toimivia ihmisiä (Paunonen 1995 [1982]; Mielikäinen 1980; 1981b; 1981c; 1986; Jonninen-Niilekselä 1982; Suojanen M. K. 1985). Projektin tuloksena saatiin paljon tietoa suomalaisen kaupunkipuhekielen äänne- ja muotovarianttien esiintymisestä sekä jonkin verran myös sanastollisesta (Jyväskylä) ja syntaktisesta (Turku) variaatiosta erilaisten aineistojen valossa.

**Variaationtutkimus**

Variaationtutkimuksen keskeisiä käsitteitä ovat *variaabeli*, *variantti*, *varieteetti* ja *variaatio*. Variantti on tutkittavan kielenpiirteen eli variaabelin ilmiasu; esimerkiksi yleiskielen *ts*:llä voi olla variantit *ts*, *tt*, *t* ja *ht* tai persoonapronominilla *minä* variantit *minä*, *mä*, *mää*, *mie* ja *miä*. *Varieteetilla* tarkoitetaan kielimuotoa (esim. Liedon murre, Helsingin puhekieli). Käsite on ongelmallinen esimerkiksi siksi, että eri kielimuotojen rajat eivät ole selväpiirteisiä ja yksilö voi käyttää useita variantteja, joista osa on murretta, osa yleiskieltä. *Variaatio* tarkoittaa kielen vaihtelua ilmiönä. Koska käsitteet ovat asultaan lähellä toisiaan, variantti ja varieteetti ovat usein sekoittuneet jopa sosiolingvistisissä tutkimuksissa.

Näissä kaikissa tutkimuksissa ja monissa muissa 1970- ja 1980-luvuilla tehdyissä sosiolingvistisissä analyyseissa perusmenetelmänä on *näennäisaikametodi*, joka Labovin (1994, 43–72) mukaan tarkoittaa sitä, että ikäryhmittäiset kielelliset erot heijastavat meneillään olevia kielenmuutoksia. Oletuksena on, että tiettynä ajankohtana tutkittujen nuorten puhetapa on heidän aikuistuessaan keski-ikäisten kielimuoto, samoin kuin keski-ikäisten kielimuoto olisi tulevien vanhusten kielimuoto. Kuten Tommi Kurki on huomauttanut, metodi ei välttämättä sovi kaikkien kielen ilmiöiden tarkasteluun. Sitä paitsi idiolektit voivat ajan oloon muuttua ja sosiaaliset muutokset muuttaa tilannetta. (Kurki 2005, 33–36.) On siten useita tekijöitä, jotka saattavat viedä kielen kehitystä aivan toiseen suuntaan kuin näennäisaikametodin avulla saatujen tulosten perusteella voisi olettaa. Varmimmin tällä metodilla saa tietoa

kulloisenkin ajankohdan synkronisesta vaihtelusta ja eri ryhmien välisistä eroista, mikä tuo esiin kieliyhteisön diversiteetin. Sen avulla voi myös hahmotella hypoteeseja mahdollisista muutoksista.

Useissa alkuvaiheen sosiolingvistisissä tutkimuksissa käsiteltiin syntyperäisten kieltä. 1980-luvulla Pohjoismaissa kiinnostuttiin maassa muuttaneiden kielestä, ja 1980-luvun alkupuolella Suomessa, Ruotsissa, Norjassa ja Tanskassa toteutettiin hanke Urbanisering och språkförändring i Norden, jossa tutkimuksen kohteena oli maalta kaupunkiin muuttaneiden kielimuoto. Keskeisenä tavoitteena oli selvittää, miten muuttajat säilyttävät lapsuuden kielimuotoaan ja miten he sopeutuvat uuden ympäristön varioiviin kielimuotoihin ja omaksuvat niiden variantteja. Suomalaiset tutkijat analysoivat Helsinkiin muuttaneita suomenkielisiä (Nuolijärvi 1986a; 1986b) ja Suomesta Eskilstunaan muuttaneita ruotsinkielisiä (Ivars 1986), Ruotsissa tutkittiin niin ikään Eskilstunaan muuttaneiden ruotsinsuomalaisten kielimuotoa (Lainio 1989). Samoin Norjassa ja Tanskassa tarkasteltiin maalta kaupunkiin muuttaneiden murteen säilymistä ja uusien varianttien omaksumista (Omdal 1994; Pedersen 1986). Kokonaiskuvan projektista saa Bengt Nordbergin (1994) toimittamasta artikkelikokoelmasta.

**Pirkko Nuolijärvi: Kenttätyötä ja variaationtutkimusta 1980-luvulla**

1980-luvun alussa käynnistyi pohjoismainen tutkimushanke Urbanisering och språkförändring i Norden, jossa tutkittiin maalta kaupunkiin muuttaneiden kielimuotoa. Tässä hankkeessa syntyi myös ensimmäinen fennistiikan sosiolingvistinen väitöskirja, tutkimukseni *Kolmannen sukupolven kieli. Helsinkiin muuttaneiden suurten ikäluokkien eteläpohjalaisten ja pohjoissavolaisten kielellinen sopeutuminen* (Nuolijärvi 1986b), ja sen sivutuotteena teos *Kieliyhteisön vaihto ja muuttajan identiteetti* (Nuolijärvi 1986a), jossa käsiteltiin muuttajien kokemuksia ja näkemyksiä kielimuodoista. Hanketta johti professori Bengt Nordberg Uppsalan yliopistosta. Suomessa tutkimusta rahoitti Suomen Akatemia.

Työ käynnistyi vuoden 1982 alussa yhteisissä kokouksissa, joissa täsmennettiin tavoitteita ja esiteltiin kunkin maan edustajien suunnitelmia.

Ennen aineistonkeruun käynnistämistä perehdyin sosiolingvistiseen kirjallisuuteen ja menetelmiin. Eteläpohjalaiset ja pohjoissavolaiset suurten ikäluokkien edustajat valikoituivat haastateltaviksi kahdesta syystä: Ryhmät edustivat hyvin erilaisia lähtömurteita, joihin julkisessa keskustelussa myös suhtauduttiin tuolloin eri tavoin. Toinen syy informanttien valintaan oli se, että suurelta osin juuri nämä ikäluokat toteuttivat niin sanotun suuren muuton eli muuttoliikkeen muualta Suomesta Etelä-Suomeen 1960-luvun lopulla ja 1970-luvun alussa. Haastateltavat etsin Väestörekisterikeskuksen tiedostojen avulla. Poimin rekisteristä henkilöitä, jotka olivat syntyneet vuosina 1943–1950 eteläpohjalaisen ja pohjoissavolaisen murrealueen kunnissa ja jotka asuivat Kalliossa, Alppilassa, Hermannissa, Vallilassa, Maunulassa ja Oulunkylässä. Sen jälkeen otin yhteyttä näihin henkilöihin, yleensä puhelimitse, mutta joskus astelin suoraan kotiovelle ja kysyin haastattelumahdollisuutta. Minulla oli paperilla lyhyt selvitys, mihin tarkoitukseen aineisto kerätään, mutta lupa kysyttiin tuolloin ainoastaan suullisesti. Vain yksi tavoittamani henkilö kieltäytyi.

Teemahaastattelut tein haastateltavien kotona kelanauhurilla maalis- ja huhtikuussa 1982, ja ne kestivät tunnin (kysymyssarja ks. Nuolijärvi 1986a; 1986b). Kaikkiaan haastatteluja oli 48, 24 kummaltakin murrealueelta, miehiä ja naisia yhtä paljon. Henkilötiedot keräsin lomakkeelle. Samaan aikaan kun haastattelin iltaisin, litteroin aineistoa päivisin. Siten sain heti tuntumaa informanttieni kielimuotoon ja pystyin jo suunnittelemaan, mihin piirteisiin kannattaa kiinnittää huomiota. Oli hyödyllistä kerätä aineisto melko lyhyessä ajassa, jotta pääsi ripeästi käsiksi analyysiin. Aineiston keruu ei ollut pelkkää materiaalin kokoamista, vaan sain eläytyä monenlaisiin elämäntarinoihin ja oppia paljon siitä, mitä muuttaminen merkitsi. Kesä- ja heinäkuussa nauhoitin vielä 12:ta informanttia heidän lapsuutensa ympäristössä, kun he keskustelivat isän, äidin tai ystävän kanssa. Näin sain aineistoa sekä kaupunki- että lomasuomesta eli aineistoa tilanteisen vaihtelun analyysin tarpeisiin. Aineiston keruu merkitsi siis lukuisia raitiovaunu- ja bussimatkoja melko painavan kelanauhurin kanssa ja kesällä juna- ja automatkoja Etelä-Pohjanmaalle ja Pohjois-Savoon.

Analysoin noin 20 äänne- ja muotopiirteen vaihtelua koko aineistossa. Koska käytin tietokonetta vasta työn viimeistelyvaiheessa, varianttien poiminta oli manuaalista, piirre kerrallaan, yksilö kerrallaan. Rakensin kustakin variaabelista taulukoita ja etsin selityksiä varianttien jakautumiselle. Tämä vaihe sisälsi siis paljon nauhojen kuuntelua ja litteraattien läpikäyntiä sekä myöhemmin sen miettimistä, miten parhaiten havainnollistaa ryhmien ja yksilöiden välisiä eroja. Rakensin yksilöprofiileja ja ilokseni sain myöhemmin huomata, että yksilöihin keskittyminen ei ollut väärä valinta; siitä tuli myöhemmin itsestään selvä osa sosiolingvistisen tutkimuksen kaanonia. Pystyin äänne- ja muotovariaation perusteella erottamaan neljä erilaista yksilöprofiilia: murteenpuhuja, leimautumaton sopeutuja eli yleispuhekielisen valtakulttuurin suosija, kaupunkilaistunut sopeutuja eli uuden alakulttuurin omaksuja ja huoliteleva puhekielestä irrottautuja. Luonnollisesti keskeistä oli myös kahdelta murrealueelta tulevan ryhmän vertailu sekä miesten ja naisten valintojen vertailu.

Pohjoismainen projektiryhmä tapasi säännöllisesti, ja vuorovaikutus tapahtui skandinaavisilla kielillä. Nämä tapaamiset olivat tärkeitä ja opettavaisia. Sen lisäksi, että oma työ sai aina uutta voimaa, opin tieteellistä vuorovaikutusta yhteisössä, jonka jäsenet tulivat eri maista, eri yliopistoista ja erilaisista keskustelukulttuureista. Projekti oli hedelmällinen: sen tuloksena syntyi kaikkiaan seitsemän väitöskirjaa ja englanninkielinen kokoomateos, jossa tuloksia esitellään artikkelimuodossa (Nordberg 1994).

Koko väitöskirjaprosessin aikana toimin aktiivisesti paikallispolitiikassa, mikä vaikutti osaltaan siihen, että labovilaisen variaationtutkimuksen perinteeseen kiinnittyvä väitöstutkimukseni oli yhteiskunnallisestikin kantaa ottavampi kuin fennistiikassa oli siihen mennessä totuttu. Elinkeino- ja aluerakenteeltaan muuttuneessa Suomessa syntyivät erilaiset kielen markkinat, joilla jokaisen oli lunastettava oma paikkansa, ja tämän yritin välittää samalla, kun analysoin yksityiskohtaisesti yksilöiden ja ryhmien kielenkäyttöä. Tutkija ei ole irrallaan muusta elämästä, ja siksi jokaisen kannattaa mennä sitä kohti, minkä tuntee omimmaksi alueekseen. Eikä kannanottojakaan kannata pelätä.

1990-luvun alussa käynnistettiin Kotimaisten kielten tutkimuskeskuksen koordinoima ja yliopistojen kanssa yhteistyössä toteutettu murteenseuruuhanke. Kymmenellä Suomen maaseutupaikkakunnalla tarkasteltiin vähintään kolmen sukupolven kielimuotoa. Paikkakunnat sijaitsevat eri puolilla Suomea, ja tavoitteena olikin saada selville, mihin suuntaan kielimuodot ovat kehittyneet eri ympäristöissä. (Kokoavat tulokset, ks. Nuolijärvi & Sorjonen 2005.) 2000-luvun alussa tehtiin uusi haastattelukierros, ja tätä aineistoa tutkittiin myös reaaliaikametodilla (esim. Kurki 2005).

*Reaaliaikaisella näkökulmalla* tarkoitetaan sitä, että samaa puhujayhteisöä tutkitaan kahtena tai useampana eri ajankohtana. Tyypillisimmin tutkimuskertojen väli on ollut 10–20 vuotta mutta joskus lyhyempikin. Reaaliaikaista tutkimusta on kahta eri tyyppiä: *paneelitutkimuksella* tarkoitetaan sitä, että tutkittavana ovat täsmälleen samat puhujayksilöt, kun taas *trenditutkimuksessa* tutkittavat valitaan kaikilla aineistonkeruukerroilla samoin kielenulkoisin kriteerein, eivätkä yksilöt ole välttämättä samoja. Monissa tutkimuksissa näitä näkökulmia on yhdistelty. (Ks. esim. Paunonen, H. 1995; 1996; 2005; 2018; Kurki 2005; Palander 2005; Mustanoja 2011; Sundgren 2002; Wagner & Buchstaller 2018).

Trendi- ja paneelitutkimusten etu verrattuna näennäisaikatutkimukseen on se, että niiden avulla saadaan tietoa siitä, mihin suuntaan tarkastelun kohteena olevat kielen ilmiöt ovat muuttuneet. Haasteena on kuitenkin erottaa se, mikä on todellista muutosta, mikä vain tilanteista vaihtelua. Erityisesti 2000-luvulla on kehitetty määrällisiä menetelmiä, jotta kielen muutosta ja variaatioon vaikuttavia tekijöitä pystyttäisiin kuvaamaan mahdollisimman luotettavasti (esim. Kurki 2005; Mustanoja 2011; Kuparinen, Mustanoja, Peltonen, Santaharju & Leino 2019; Lappalainen, Mustanoja & O'Dell 2019). Tutkimukset ovat osoittaneet, etteivät idiolektit ole staattisia vaan erisuuntaiset muutokset ovat pikemminkin sääntö kuin poikkeus, myös myöhemmällä aikuisiällä. Eri yksilöillä muutokset ovat kuitenkin erisuuntaisia ja kumoavat toisensa niin, että yhteisötasolla muutos ei useinkaan ole radikaali (Mustanoja 2011). Reaaliaikatutkimusten huonona puolena on tutkimusten työläys, pitkä aikajänne ja se, että ensimmäisen vaiheen ratkaisut voivat rajoittaa toisen vaiheen ratkaisuja. Tällaisia tutkimuksia on tehty Suomessa

1990-luvulta lähtien, mutta nyt kun sosiolingvistiikalla on jo kymmenien vuosien historia ja aiempia aineistoja on kertynyt runsaasti, tällaiselle pitkittäistutkimukselle on aiempaa paremmat edellytykset.

Kaikkineen määrällisillä menetelmillä on ollut kansainvälisessä sosiolingvistiikassa keskeinen sija. Labov kehitteli tutkimusryhmineen jo 1970-luvulla Varbrul-monimuuttujamenetelmän, jolla voidaan selvittää samanaikaisesti useiden muuttujien selitysvoimaa. Menetelmä palvelee sosiolingvistiikan keskeisiä lähtökohtia yhdistäessään vaihtelun sosiaalisen ja lingvistisen ulottuvuuden, sillä analyysissa on mahdollisuus tarkastella sekä kielenulkoisia (esim. ikä, sukupuoli) että kielen sisäisiä (esim. äänneasema) muuttujia. Varbrulin pohjalta on kehitetty myöhemmin Goldvarb-niminen menetelmä. (Esim. Tagliamonte 2006.)

Suomalaisessa sosiolingvistiikassa tilastollisia menetelmiä on sovellettu vasta 2000-luvulla, ensimmäisenä Tommi Kurki (2005), joka on hyödyntänyt Varbrul-ohjelmaa tutkiessaan Alastaron murteen muutosta reaaliaikaisen aineiston valossa. Muita suomalaisessa sosiolingvistiikassa hyödynnettyjä tilastomenetelmiä ovat bayesilainen menetelmä, jossa keskeistä on todennäköisyyden tulkinta (esim. Mustanoja 2011), sekä päätöspuumenetelmä (engl. *detection tree*), joka on Varbrulin tavoin monimuuttujamenetelmä ja josta on edelleen kehitetty niin sanottu satunnaismetsämenetelmä (engl. *random forest*; ks. esim. Priiki 2017; ks. myös Määrällinen korpuslingvistiikka tk.). Tilastomenetelmien käyttö on paitsi antanut pohjaa sen tulkinnalle, millainen variaatio on tulkittavissa muutokseksi (esim. Mustanoja 2011), myös paljastanut sellaisia muuttujia, joiden merkitystä olisi vaikeaa tunnistaa muilla tavoilla. Jälkimmäisestä sopii esimerkiksi Katri Priikin väitöstutkimus (2017), jossa hän osoittaa, että kaakkoissatakuntalaisessa puheessa 3. persoonan pronominien vaihtelua keskeisesti säätelevä tekijä on virittyneisyys (engl. *priming*) eli se, mitä pronominia on edellä käytetty samasta tarkoitteesta.

## 3.2. Kielenvaihto ja kielen säilyttäminen kielisosiologian näkökulmasta

Erityisesti kielivähemmistöt ja alkuperäiskansat, joiden kielellä on vähemmän puhujia ja heikompi juridinen asema kuin enemmistökielellä, joutuvat jatkuvasti kamppailemaan oman kielensä säilymisen puolesta. Kun virallisessa vuorovaikutuksessa ja kodin ulkopuolella käytetään hallitsevan väestön kieltä, jää oman kielen käyttöala vähäiseksi ja kielen säilyminen on uhanalainen. Eri kieliryhmien tilannetta Euroopan maissa on tutkittu ELDIA-hankkeessa (European Language Diversity for All) ja kehitetty kielibarometri tilanteen kuvaamiseksi ja seuraamiseksi (Laakso, Sarhimaa, Spiliopoulou Åkermark & Toivanen 2016; ks. myös Sarhimaa 2017). Alkuperäiskansan tai vähemmistökielisten olot voivat nyky-yhteiskunnassa muuttua suurestikin, kuten Helsingin saamelaisilla (Lindgren 2000) tai Pietarin seudun inkeriläisillä, jotka elävät venäjänkielisen yhteisön keskellä (Mononen 2013).

Kielenvaihdon problematiikkaa ja edellytyksiä on käsitelty nimenomaan kielisosiologisesta näkökulmasta (varhaisista tutkimuksista ks. esim. Fishman 1972). Usein tutkimus on kiinnittänyt huomionsa vain yhteen aikatasoon ja tarkastellut kielen tilannetta synkronisesti. Kielenvaihto ei kuitenkaan tapahdu hetkessä, vaan se voi toteutua usean sukupolven aikana. Kaksikielisyyden tutkija Einar Haugen luonnehti aikanaan kielenvaihdon prosessia kielenvaihdon kierteeksi: Yksikielisestä vähemmistökielisestä yksilöstä tai ryhmästä tulee vähitellen kaksikielinen, kun kontakti toisenkieliseen yhteisöön on jatkuvaa. Ajan oloon alun perin vieras kieli valtaa sitten sijaa yksilön tai ryhmän elämässä, ja kielenvaihto on alkanut. Lopulta alkuperäisen kielen käyttö vähenee ja äärimmäisessä tapauksessa loppuu kokonaan. (Haugen 1980.) Samassakaan yhteisössä kaikille yksilöille ei kuitenkaan käy samalla tavoin, vaan osa yksilöistä voi säilyttää kielensä. Haugenin kuvaus ei kuitenkaan oikein vastaa todellisuutta, koska vähemmistökieliset yksilöt vain harvoin ovat yksikielisiä. Sen sijaan se havainto, että yksilöiden kielellinen historia on erilainen, pitää paikkansa.

Suomessa kielenvaihdon kierrettä ovat tutkineet muiden muassa Marjut Aikio ja Marika Tandefelt. Muualle muuttaneiden suomenkielisten

kielenvaihtoa on selvitetty australiansuomalaisten keskuudessa (ks. esim. Kovács 2004), ja suomenkielisten vähemmistöjen tilannetta on kuvattu myös monissa muissa ympäristöissä (ks. esim. Jönsson-Korhola & Lindgren 2003).

Aikio (1988) on käsitellyt viiden saamelaiskylän kielenvaihtoa vuosina 1910–1980 ja Tandefelt (1988) ruotsinkielisen suvun kolmen sukupolven kielenvaihtoa ja kielen säilyttämistä Vantaalla (Helsingin pitäjässä). Aikio kehitti työssään kolmenlaisia indeksejä, joiden avulla hän tarkasteli ja vertaili informanttiryhmiään: perusindeksinä on kielenkäyttöindeksi, ja muut kaksi indeksiä ovat kielen ymmärtämisindeksi ja kielitaitoindeksi. Kielenvalintaa Aikio tarkasteli sisarussarjoittain, kylittäin, sukupolvittain ja perheittäin. Lisäksi hän valitsi lähemmän tarkastelun kohteeksi muutamia avaininformantteja.

Tandefeltin tutkimus on sosiolingvistisempi, ja se käsittelee variaationtutkimusten tapaan joukkoa kielenpiirteitä, joiden kehitystä tarkastellaan siltä kannalta, miten murre muuttuu standardikielisemmäksi. Toinen näkökulma koskee sitä, miten kieli vaihtuu ruotsista suomeksi. Vantaalaisympäristössä saman, alkuaan yksikielisen suvun jäsenet erkaantuvat tulosten mukaan kahteen suuntaan: osa vaihtaa kielensä, osa säilyttää ruotsin suomen kielen rinnalla. Tandefeltin tutkimus pohjautuu Einar Haugenin (1971) ajatuksiin kielen säilyttämiseen vaikuttavista tekijöistä.

**Kielenvaihto suvuissa**

Eri sukuhaarojen kielenvaihtoa tai kielen säilymistä voi kuvata, kuten Tandefelt väitöstutkimuksessaan (1988), esimerkiksi seuraavin symbolein[1]:

| Sukupolvi | A | B | C | D | E |
|---|---|---|---|---|---|
| 1) | Ab > | Ab > | AB > | AB > | (AB) |
| 2) | Ab > | AB > | AB > | AB > | (AB) |
| 3) | Ab > | Ab > | AB > | BA > | B |
| 4) | Ab > | Ab > | AB > | Ba > | B |
| 5) | Ab > | Ab > | AB > | B > | B |

Sukupolven A pääasiallinen kieli on ruotsi (A). Sukupolvessa B on yksi yksilö (2), jolla kielet ovat yhtä vahvoja. Sukupolvessa C kielten tasavahva asema koskee kaikkia puhujia (1–5). Sukupolvessa D näkyy eriytyminen: osalla kielet ovat yhtä vahvoja, mutta suomi (B) voi olla ensisijainen (esim. yksilöt 3 ja 4), jolloin ruotsi jää vähemmälle käytölle (a) tai sitä ei enää käytetä (yksilö 5). Sukupolvessa E suomi on vallannut tilan ruotsilta tai molempia käytetään yhtä paljon.

Mainitun kaltaiset tutkimukset ovat metodeiltaan rikkaita. Kielisosiologinen tarkastelu vaatii sekä tutkittavan ympäristön ja sen historian tarkkaa tuntemusta että sosiolingvististen ja kielisosiologisten menetelmien yhdistelyä ja kehittämistä kulloisenkin analyysin tarpeisiin. Nykyoloissa esimerkiksi sukupolvi sitten maahan tulleiden monikielisten yksilöiden kielenvaihto voi olla vielä nopeampaa kuin Aikion ja Tandefeltin tutkimien ryhmien ja yksilöiden kielenvaihto. Tämä vaatii yhä enemmän menetelmien kehittämistä ja asianomaisten kieliryhmien olojen tuntemusta.

## 3.3. Variaatio ja vuorovaikutus

2000-luvulta lähtien suomalaista variaationtutkimusta on rikastanut myös vuorovaikutuksen näkökulma. Ensi kertaa se oli mukana Pirkko Nuolijärven ja Liisa Tiittulan (2000) poliittisia televisiokeskusteluja koskevassa tutkimuksessa, jossa variaatiota ja vuorovaikutusta tarkasteltiin selvästi erikseen mutta kuitenkin samasta aineistosta ja saman tutkimuksen puitteissa. Selvemmin niitä yhdistettiin Hanna Lappalaisen väitöskirjassa (2004), jossa variaation funktioille etsitään selityksiä vuorovaikutuskontekstia analysoimalla. Pyrkimyksenä on nähdä numeroiden taakse ja tutkia, mitkä muut kuin kielenulkoiset tekijät voivat selittää kielellistä vaihtelua. Lappalaisen tutkimuksessa muun muassa ilmeni, että koodinvaihdoilla arkisesta puhekielestä hyperkorrektiin yleiskieleen oli tutkitussa ryhmässä selviä vuorovaikutustehtäviä ja niillä

esimerkiksi osoitettiin toiseen rooliin tai toimintoon siirtymistä. Metodisesti tutkimuksessa nojaudutaan keskustelunanalyysiin (ks. myös esim. Lappalainen 2006; 2011; keskustelunanalyysista ks. Keskustelunanalyysi ja vuorovaikutuslingvistiikka tk.).

**Koodinvaihto**

Lappalainen (2004) on osoittanut väitöskirjassaan, miten keskustelijat voivat siirtymällä tavanomaisesta puhekielisestä ilmaisutavastaan korosteisen yleiskieliseen puhetapaan ilmentää monenlaisia kontrasteja ja osoittaa vuorovaikutustilanteessa tapahtuvaa muutosta. Esimerkiksi kun lapsi temppuilee ruokapöydässä vieraiden läsnä ollessa eikä äiti saa häntä tottelemaan, isän huolitellun yleiskielisesti esittämän lausuman ”mitä vielä keksimme sinua uhkaillessa” voi tulkita yritykseksi keventää tilannetta ja käsitellä sitä humoristisesti. Äiti ei kuitenkaan tähän tarjoukseen tartu vaan jatkaa samassa vakavassa moodissa, ja näin konflikti jatkuu. Tilanteissa, joissa muutkin osallistujat kuin koodinvaihtoa käyttänyt puhuja siirtyvät samaan puhetapaan, yleiskielen käyttö voi toimia yhteisyyden rakentamisen keinona.

2010-luvulla variaation ja vuorovaikutuksen näkökulmia yhdistävät tutkijat ovat nimenneet tutkimuksensa vuorovaikutussosiolingvistiikaksi (ks. esim. Mononen 2013; Lehtonen 2015; Rouhikoski 2015; Priiki 2017). Sille on ominaista tutkia, mitä variaatiolla tehdään vuorovaikutuksessa ja millaisia merkityksiä vuorovaikutustilanteen osallistujat variaatiolle antavat. Tutkimuskohteena ovat olleet niin arkiset kuin institutionaaliset tilanteet ja sekä yksi- että monikielinen vuorovaikutus. Tutkittavat ilmiöt ovat monipuolistuneet jo osin siksi, että haastattelujen lisäksi aineistona ovat olleet muunkinlaiset vuorovaikutustilanteet. Vuorovaikutussosiolingvistiikka on tehnyt näkyväksi sen, että eri puhetilanteiden vertailun lisäksi on tarkasteltava vaihtelua myös vuorovaikutustilanteiden sisällä, mikrotasolla, samalla kuitenkin pohtien sen kytköksiä makrotasolle. Vaihtelun selittämisessä on hyödynnetty esimerkiksi *diskurssi-identiteetin*, *osallistumiskehikon* (engl. *participation framework*) ja *asennoitumisen* (engl. *stance*) käsitteitä.

## 3.4. Variaatioon liittyvät käsitykset, asenteet ja ideologiat

Kieliasenteista ja -käsityksistä oltiin kiinnostuneita jo suomalaisen sosiolingvistiikan alkuvaiheissa (esim. Mielikäinen 1981a; Nuolijärvi 1986a), mutta laajamittaisemman ja systemaattisemman tutkimuksen kohteeksi ne ovat nousseet tällä vuosituhannella. Tähän on vaikuttanut erityisesti kansanlingvistisen tutkimuksen rantautuminen Suomeen. Kansanlingvistiikassa lähtökohtana on tavallisten kielenkäyttäjien kielikäsitysten ja -asenteiden tarkastelu. Niitä tutkitaan erilaisin menetelmin, kuten kuuntelutesteillä (esim. murrenäytteiden paikantaminen), kysymällä suoraan eri kielimuotoihin liittyvistä arvotuksista, piirrättämällä murrekarttoja ja pyytämällä tutkittavia nimeämään murrealueita ja -piirteitä ja imitoimaan jotain kielimuotoa. Näillä menetelmillä on saatu tietoa esimerkiksi siitä, millaisiin kielenpiirteisiin kielenkäyttäjät kiinnittävät omassa murteessaan ja muiden murteissa huomiota, miten murrealueita hahmotetaan maantieteellisesti ja mitä murteita pidetään kauneimpina. Keskeinen tutkimuskohde on myös metakieli eli se, miten kielenkäyttäjät eri kielenilmiöistä puhuvat, kun heillä ei luonnollisestikaan voi olla samoja käsitteitä hallussaan kuin ammatikseen kieltä analysoivilla tutkijoilla. (Preston 1989; 1999; Long & Preston 2000; Vaattovaara 2009; Palander 2011; Mielikäinen & Palander 2014; Kurki 2018; Vaattovaara, Kunnas & Saviniemi 2018.) Kielenpiirteiden variaatiota, leimautumista ja nykysuomen murteiden jäsentymistä on tarkasteltu myös kielellisen identiteetin rakentumisen näkökulmasta (Mantila 2004).

**Kieliasenteet**

Johanna Vaattovaara on väitöstutkimuksessaan (2009) osoittanut, että kieliasenteiden dynamiikka kuuluu variaationtutkimuksen ytimeen eikä ole siitä irrallinen ilmiö. Kielitietoisuuden analyysi yhdistettynä variaationanalyysiin voi murtaa annettujen sosiolingvististen kategorioiden (ikä, sukupuoli, koulutus yms.) luomia stereotypioita kuten Tornionlaaksossa: murteelle ominaisen *h*:n käytöllä ei rakenneta maskulii-

nisuuden merkityksiä, kuten aiemmin on oletettu, vaan sillä viestitään sitoutumista kotiseutuun, mikä on yleisempää nuorten miesten kuin nuorten naisten keskuudessa.

Kansanlingvistisen tutkimuksen tavoitteena on päästä käsiksi kielenkäyttäjien kielelliseen tietoisuuteen ja sitä kautta saada myös ymmärrystä kielen muuttumiseen vaikuttavista tekijöistä. Taustalla on vahvasti spatiaalinen ajattelu, jossa kielen oletetaan olevan kytköksissä paikkaan. Vaikka kielen ja paikan suhde on monin tavoin olennainen, tätä näkemystä on myös kritisoitu turhan kapeana: kieleen kytkeytyvät mielikuvat voivat olla sidoksissa jopa paljon enemmän johonkin muuhun ulottuvuuteen kuin paikkaan. Yksi esimerkki tästä on niin sanottu stadilainen ässä, johon liittyy paljon muitakin sosiaalisia merkityksiä kuin kytkös pääkaupunkiseutuun. (Vaattovaara 2014; Vaattovaara & Halonen 2015.) Vuorovaikutuksen, asenteiden ja ideologioiden erittely onkin tärkeä osa kielentutkijoiden yhteiskunnallista vaikuttamista ja kielenkäyttäjien kielitietoisuuden vahvistamista (ks. esim. Piippo, Vaattovaara & Voutilainen 2016). Kansanlingvistisen tutkimuksen lisäksi Suomessa on tehty muunlaistakin tutkimusta, jolla on pyritty selvittämään kieliasenteita ja -käsityksiä (esim. Peterson & Vaattovaara 2014). Yksi keskeisistä tutkimusmenetelmistä on ollut verkossa toteutettu kyselytutkimus (ks. Kysely tk.).

Suhde standardikieleen sekä sen ja muiden kielimuotojen varianttien vaihtelu on alusta pitäen ollut sosiolingvistisen tutkimuksen kohteena – samoin kuin normit ja niiden luonne (ks. esim. Piippo 2012). Tietoisuus standardikielen piirteistä ja mielipiteet sen ja murteiden suhteesta ovat myös asennetutkimusten keskeisiä sisältöjä. 1990-luvulta lähtien on sosiolingvistiikassa kiinnitetty runsaasti huomiota siihen, että nykyisissä kieliyhteisöissä on meneillään kehityskulkuja, jotka asettavat *standardikielen* käsitteen uuteen valoon. Kysymys on siitä, miten kieliasenteet muuttuvat, miten median vaikutus näkyy kielenkäyttäjien valinnoissa ja miten puhutun kielen piirteet vahvistuvat myös niin sanotussa korrektissa kielenkäytössä (engl. ”*best language*”) samoin kuin erilaisissa kirjoitetuissa tekstilajeissa. Näitä seikkoja on tarkasteltu lukuisissa tutkimuksissa. Yhtenä esimerkkinä mainittakoon eurooppalainen

SLICE-ohjelma (Standard Language Ideology in Contemporary Europe), joka on käsitellyt eri maiden kieliyhteisöjen kehitystä tästä näkökulmasta. Keskeisiä käsitteitä tämän ohjelman tutkimuksissa ovat *destandardisaatio* ja *demotisaatio* (ks. esim. Kristiansen & Coupland 2011; Maegaard, Gregersen, Quist & Jørgensen 2009). *Destandardisaatio* merkitsee etääntymistä standardikielestä, *demotisaatio* puolestaan tarkoittaa sitä, että standardikieltä käytetään entistä laajemmin, ilman että sen status tai kieli-ideologiat muuttuvat. Demotisaatiosta on kyse esimerkiksi silloin, kun standardi korvaa murteet niiden aiemmissa käyttöyhteyksissä.

Kuvaava esimerkki standardin muuttumisesta on Tanskasta, jossa Kööpenhaminassa käytetyt variantit ovat asettuneet osaksi muualla elävien nuorten standardia. Kysymys on eräänlaisesta kööpenhaminalaisvarieteetin keskiverrosta, ei kaikkein leimallisimmasta pääkaupunkilaispuheesta (Kristiansen 2000). Suomessakin on nähtävissä samantapaisia ilmiöitä; näitä ei kuitenkaan ole käsitelty erityisesti destandardisaation näkökulmasta vaan osana paikallisten kielimuotojen muutoksen kuvausta (ks. kuitenkin Nuolijärvi & Vaattovaara 2011).

## 3.5. Kieli ja vuorovaikutus monikielisissä toimintayhteisöissä

Suomi on ollut kautta aikain monikielinen maa, ja eri kieliä on käytetty rinnakkain ja lomittain. Yksi kiintoisa esimerkki tästä on Stadin slangi, joka on syntynyt suomen ja ruotsin kohtaamisesta (ks. esim. Paunonen, H. & Paunonen, M. 2000). Viime vuosikymmenten muuttoliikkeessä uutta on kuitenkin se, että varsin lyhyessä ajassa maahamme on tullut huomattava määrä kymmenien eri kielten puhujia ja varsinkin pääkaupunkiseudulla on alueita, joilla maahanmuuttajataustaisten osuus on suuri. Erityisesti tämä näkyy kouluissa, joihin on muodostunut monikielisiä toimintayhteisöjä. Näissä yhteisöissä eritaustaiset osallistujat voivat jakaa kielellisiä resurssejaan ja hyödyntää vuorovaikutuksessaan sellaisia kieliä, joita eivät varsinaisesti osaa. Tällaista kielellisten resurssien hyödyntämistä kutsutaan *kielenylitykseksi*. Päästäkseen selville erilaisiin kielellisiin resursseihin linkittyvistä sosiaalisista merkityksistä

tutkijan on usein vietettävä runsaasti aikaa tutkimassaan yhteisössä. Etnografisella tutkimusotteella onkin keskeinen sija tämäntyyppisessä monikielisyyden tutkimuksessa. (Rampton 1995; 2006; Lehtonen 2015: ks. myös Lingvistinen etnografia toimintayhteisöissä tk.)

Monikielisten yhteisöjen tutkimuksessa kieltä on tarkasteltu toimintana, ei staattisena oliona, ja siinä on irtaannuttu *varieteetin* käsitteestä, jota on kyseenalaistettu myös jo aiemmassa sosiolingvistisessä keskustelussa (ks. esim. Nuolijärvi 2001). Monikielisten ryhmien analyysissa toimintayhteisön vuorovaikutus ja kielellisten taitojen hyödyntäminen pääsevät esille systemaattisesti, kun tutkitaan kahta tai useampaa kieltä sujuvasti käyttävien puhujien välistä vuorovaikutusta. Kysymys on jälleen kielen muutoksen tutkimuksesta: mitä tapahtuu, kun eri kielten puhujat ovat vuorovaikutuksessa samassa ympäristössä, kuten koululuokassa tai harrastuspiireissä?

**Sosiaalinen indeksi**

Eri kielten puhujien läheinen vuorovaikutus voi näkyä hyvin konkreettisesti, kuten Heini Lehtonen osoittaa väitöstutkimuksessaan (2015). Monietnisissä nuorisoryhmissä kiinteästi liikkuvat helsinkiläiskoululaiset käyttävät kielitaustastaan riippumatta ilmausta *wallahi* '(vannon) Allahin nimeen' ja sen suomenkielisiä vastineita *mä vannon* ja *mä lupaan* samaan tapaan kuin vastaavia ilmauksia käytetään monissa muissakin maissa. Lehtosen mukaan ne toimivat sosiaalisina indekseinä hänen tutkimiensa nuorten vuorovaikutuksessa.

Monikielisyys ei ole siten vain eri kielten olemassaoloa kielenkäyttäjien elämässä vaan toimintaa monenlaisilla kielellisillä resursseilla ja niiden vaikutusta toisiinsa arjen vuorovaikutustilanteissa. Ilmiö ei sinänsä ole mitenkään uusi, mutta Suomen oloissa muutos näkyy siten, että yhä uudet kielet vaikuttavat myös suomenkielisten kielenkäyttöön. Toimintayhteisöjen etnografinen tutkimus kiinnittää huomiota erilaisten kielellisten resurssien yhdistämiseen ja sen funktioihin eri tilanteissa. Tätä

suuntausta voi nimittää myös *globalisaation sosiolingvistiikaksi* (Blommaert 2010; ks. myös Lehtonen 2015, 27). Sen keskeinen käsite on *superdiversiteetti*; yhdessä yhteisössä käytetyt kielet eivät ole erillisiä eivätkä kielten rajat selviä, vaan yksilöiden kielelliset repertuaarit koostuvat erilaisista resursseista, joita voidaan hyödyntää joustavasti samassa tilanteessa. Kun tietyt resurssit kimputtuvat ja saavat sosiaalisia merkityksiä eli sosiaalista indeksisyyttä, ne muodostavat tyylejä – tyypillisesti yhdessä muiden semioottisten resurssien kanssa. Prosessia, jossa tällaiset sosiaalisesti tunnistettavat tyylit tai rekisterit syntyvät ja muokkautuvat, kutsutaan *rekisteriytymiseksi*. Rekisteriytyminen on jatkuva prosessi, vaikka se kiinnittyy aina myös aikaan ja paikkaan. (Agha 2007; ks. Lehtonen 2015, 35–50.)

# 4. Yhteenveto

Tässä artikkelissa olemme esitelleet erityisesti niitä sosiolingvistiikan suuntauksia, jotka 1970-luvulta alkaen ovat olleet edustettuina suomen kielen tutkimuksessa. Suuntauksista näennäisaika- ja reaaliaikametodia käyttävä, äänne- ja muotopiirteitä tarkasteleva variaationanalyysi on keskeisessä osassa. Myös sosiolingvistisen variaationanalyysin yhdistäminen vuorovaikutuksen tutkimuksen menetelmiin on ollut suhteellisen yleistä. Se selittyy osin siitä, että Suomessa on vahva, erityisesti keskustelunanalyysin menetelmin harjoitettu vuorovaikutuksen tutkimus, joka on ollut omiaan tukemaan myös sosiolingvistien työtä. Niin ikään kielellisten asenteiden tutkimus on saanut jalansijaa, ja kiinnostuksen kohteena ovat olleet erityisesti käsitykset eri murteista ja puhutun kielen muodoista sekä niiden hierarkiasta. 2010-luvulla on lisääntynyt myös monikielisten yhteisöjen etnografinen tutkimus, joka kohdistaa huomionsa erikielisten, samassa yhteisössä elävien ihmisten kielenkäytön ilmiöihin ja siihen, miten eri kielet vaikuttavat toisiinsa. Selvästi kielisosiologiset lähestymistavat ovat jääneet viime aikoina vähemmälle, mutta uudenlaisen monikielisyyden keskellä kielisosiologinen ote tulee varmasti vahvistumaan.

Menetelmien ja aineistojen monipuolistuminen avaa yhä uusia näkymiä sosiolingvistiselle tutkimukselle. Kuten sosiolingvistit Jan Blommaert ja Ben Rampton (2011) ovat huomauttaneet, sosiolingvistiikan tulee reagoida yhteiskunnan muutoksiin. Keskiöön on nostettava liikkuvuus, sekoittuminen, poliittinen dynamiikka ja historiallinen kehys siten, kuin ne ilmenevät kielissä ja vuorovaikutustilanteissa nyky-yhteiskunnassa.

Aineistoja on mahdollista koota nykyisin hyvin monista eri lähteistä; laajoja kirjoitetun ja puhutun kielen korpuksia on helposti saatavilla enemmän kuin koskaan. Tällaisia korpuksia ovat esimerkiksi Nykysuomalaisen puhekielen murroksen aineisto, Helsingin puhekielen pitkittäiskorpus, Kotimaisten kielten keskuksen Murteenseuruukorpus ja uusimpana Sata suomalaista kielellistä elämäkertaa -korpus (ks. Priiki, Mononen & Lappalainen 2018), jotka kaikki ovat Kielipankissa tai tulossa Kielipankkiin. Siten esimerkiksi tutkimuksen fokusoiminen yhteen piirteeseen onnistuu eri aineistoja ja näkökulmia hyödyntäen. Laajat aineistot myös luovat entistä paremman perustan kvantitatiivisten menetelmien käytölle ja kehittämiselle. Eri kielentutkimuksen metodeja yhdistellään todennäköisesti yhä enemmän, ja eri alojen rajat voivat siten madaltua. Yksittäisten kielellisten muuttujien sijasta sosiolingvistiikassakin keskitytään enemmän vuorovaikutuksen tyyleihin ja etsitään kielellisille eroille selityksiä elämäntavoista ja tilanteisista seikoista. Koska sosiolingvististä analyysia on tehty jo useiden vuosikymmenien ajan ja aineistoa on tarjolla eri aikakausilta, on hyvät lähtökohdat myös reaaliaikaiselle tutkimukselle, jonka ansiosta on mahdollista saada entistä tarkempi kuva kielen kehityksestä. Yhä enemmän on tarvetta yhdistää lingvististä analyysia muihin tieteenaloihin, esimerkiksi yhteiskuntatieteisiin, koska on ilmeistä, että kielenkäytön luokka-analyysille on uudenlainen tilaus eriarvoisessa tiimalasiyhteiskunnassa, jossa rikkaiden ja köyhien elämät ja vuorovaikutus erkanevat toisistaan entistä leimallisemmin myös Suomessa.

Sosiolingvistiikan tuloksia voi soveltaa monissa yhteyksissä. Kielen ja vuorovaikutuksen variaatio koskettaa kaikkia yhteiskunnan tasoja. Sosiolingvistit voivat työllään lisätä päiväkotien ja koulujen henkilöstön tietämystä erilaisista elämäntyyleistä ja puhetavoista, viranomaisten ja median viestintä hyötyy eri aineistoilla tehdyistä tutkimuksista,

ja yritysten viestintä tarvitsee tietoa monentasoisesta variaatiosta ja kielenkäytön muutoksista. Tämä kaikki edellyttää tutkimuksen jatkuvaa yleistajuistamista ja aktiivista toimintaa myös tiedeyhteisön ulkopuolella. Hyvänä esimerkkinä tästä on yhteisölähtöinen toimintatutkimus ”Itä-Helsingin uudet Suomen kielet”, jossa tutkijat tekevät näkyväksi tutkimassaan yhteisössä ilmenevää monikielisyyttä ja luovat käytäntöjä, joilla tätä erilaisten kielten taitoa voitaisiin yleisemminkin hyödyntää paremmin suomalaisessa yhteiskunnassa (ks. Lehtonen & Räty 2018). Sosiolingvistiikka on ylipäätään niitä kielentutkimuksen alueita, joiden tuloksia on suhteellisen helppo välittää esimerkiksi opetuksessa ja näin vahvistaa eri ryhmien ja yksilöiden kielitietoisuutta.

**Aiheesta lisää:**


Bayley, Robert, Cameron, Richard & Ceil, Lucas (toim.). 2013. *The Oxford Handbook of Sociolinguistics*. Oxford: Oxford University Press.

Coupland, Nikolas (toim.). 2016. *Sociolinguistics. Theoretical Debates*. Cambridge: Cambridge University Press.

Meyerhoff, Miriam. 2011. *Introducing Sociolinguistics.* (2. p.). Abingdon: Routledge.

Piippo, Irina, Vaattovaara, Johanna & Voutilainen, Eero (toim.). 2016. *Kielen taju. Vuorovaikutus, asenteet ja ideologiat*. Helsinki: Art House.


## VIITTEET

1 A–E = sukupolvet; 1–5 = yksilöt; A = ruotsi, B = suomi; suuraakkonen = vahvempi kieli puhujalla, pienaakkonen = heikompi kieli puhujalla; aakkosten järjestys = ensisijainen ja toissijainen kieli.

## KIRJALLISUUS


Agha, Asif. 2007. *Language and Social Relations*. Cambridge: Cambridge University Press.

Aikio, Marjut. 1988. *Saamelaiset kielenvaihdon kierteessä. Kielisosiologinen tutkimus viiden saamelaiskylän kielenvaihdosta 1910–1980*. SKST 479. Helsinki: Suomalaisen Kirjallisuuden Seura.

Bayley, Robert, Cameron, Richard & Ceil, Lucas (toim.). 2013. *The Oxford Handbook of Sociolinguistics*. Oxford: Oxford University Press.

Blommaert, Jan. 2010. *The Sociolinguistics of Globalization*. Cambridge approaches to language contact. Cambridge: Cambridge University Press.

Blommaert, Jan & Rampton, Ben. 2011. Language and superdiversity: A position paper. *Working papers in urban language & literacies* 70. [verkkoaineisto]. [viitattu 2.5.2018]. Saatavissa: http://www.academia.edu/6356809/WP70.

Chambers, J. K. 2009. *Sociolinguistic Theory. Linguistic Variation and Its Social Significance*. (3. p.). Oxford: Blackwell.

Coulmas, Florian. 1997. Introduction. Julkaisussa: Bayley, Robert, Cameron, Richard & Ceil, Lucas (toim.) *The Handbook of Sociolinguistics*. Oxford: Blackwell, 1–11.

Coupland, Nikolas. 2007. *Style. Language, Variation and Identity*. Cambridge: Cambridge University Press.

Coupland, Nikolas. 2009. Dialects, standards and social change. Julkaisussa: Maegaard, Marie, Gregersen, Frans, Quist, Pia & Jørgensen, J. Normann (toim.) *Language Attitudes, Standardization and Language Change*. Oslo: Novus, 27–49.

Coupland, Nikolas (toim.). 2016. *Sociolinguistics. Theoretical Debates*. Cambridge: Cambridge University Press.

Coupland, Nikolas & Jaworski, Adam (toim.). 1997. *Sociolinguistics. A Reader and Coursebook*. Houndmills: Macmillan Press.

Coulthard, Malcolm & Johnson, Alison. 2007. *An Introduction to Forensic Linguistics. Language in Evidence*. London: Routledge.

Currie, Haver C. 1952. A projection of sociolinguistics: the relationship of speech to social status. *Southern Speech Journal* 18, 28–37.

Eckert, Penelope. 2000. *Linguistic Variation as Social Practice. Linguistic Construction of Identity in Belten High*. Oxford: Blackwell.

Eckert, Penelope. 2012. Three waves of variation study. The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41, 87–100.

Eckert, Penelope & Labov, William. 2017. Phonetics, phonology and social meaning. *Journal of Sociolinguistics* 21, 467–496.

Fishman, Joshua A. 1968. *Readings in the Sociology of Language*. The Hague: Mouton.

Fishman, Joshua A. 1972. Language maintenance and language shift as a field of inquiry. Julkaisussa: Fishman, Joshua A. (toim.) *Language in Sociocultural Change*. Stanford (CA): Stanford University Press, 76–133.

Gardin, Bernard & Marcellesi, Jean-Babtiste. 1987. The subject matter of sociolinguistics. Julkaisussa: Ammon, Ulrich, Dittmar, Norbert & Mattheier, Klaus J. (toim.) *Sociolinguistics. An International Handbook of the Science of Language and Society* 1. Berlin: De Gruyter, 16–25.

Giles, Howard & St Clair, Robert N. (toim.). 1979. *Language and Social Psychology*. Oxford: Basil Blackwell.

Gregersen, Frans (toim.). 2014. *Hvor parallelt. Om parallellspråkighet på Nordens universitet.* TemaNord 535. København: Nordisk Ministerråd.

Gumperz, John J. (toim.). 1982. *Language and Social Identity.* Studies in interactional sociolinguistics 2. Cambridge: Cambridge University Press.

Haugen, Einar. 1971. The ecology of language. Julkaisussa: *The Linguistic Reporter* 13, liite 25, 19–26.

Haugen, Einar. 1980. Language problems and language planning. The Scandinavian model. Julkaisussa: Nelde, Peter Hans (toim.) *Zeitschrift für Dialektologie und Linguistik.* Beiheft 32. Sprachkontakt und Sprachkonflikt. Wiesbaden: Franz Steiner Verlag, 151–157.

Hymes, Dell. 1971. Sociolinguistics and the ethnography of speaking. Julkaisussa: Ardener, Edwin (toim.) *Social Anthropology and Language.* London: Tavistock, 47–94.

Ivars, Ann-Marie. 1986. *Från Österbotten till Sörmland. En undersökning av emigration och språklig anpassning.* Studier i nordisk filologi 66. Helsingfors: Svenska litteratursällskapet i Finland.

Jahnukainen, Raili. 1954. Havaintoja siirtoväen murteen muuttumisesta Kokemäellä. *Virittäjä* 58, 366–376.

Johnstone, Barbara. 2011. Dialect enregisterment in performance. *Journal of Sociolinguistics* 15, 657–679.

Jonninen-Niilekselä, Kaija (toim.). 1982. *Tampereen puhekieli tutkimuskohteena*. Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitoksen julkaisuja 6. Folia Fennistica & Linguistica 6. Tampere: Tampereen yliopisto.

Juusela, Kaisu. 1994. Variation as a theme in Finnish dialectology and sociolinguistics in recent years. Julkaisussa: *Fenno-Ugrica Suecana* 12. Tidskrift för finsk-ugrisk forskning i Sverige. Finska institutionen. Stockholm: Stockholms universitet, 1–54.

Jönsson-Korhola, Hannele & Lindgren, Anna-Riitta (toim.). 2003. *Monena suomi maailmalla. Suomalaisperäisiä kielivähemmistöjä*. Tietolipas 190. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kallmeyer, Werner. 1987. Stadtsprache als ein Schwerpunkt soziolinguistischer Forschung in Europa. *Sociolinguistica* 1, 80–99.

Kallmeyer, Werner. 1994. Das Projekt "Kommunikation in der Stadt". Julkaisussa: Kallmeyer, Werner (toim.) *Kommunikation in der Stadt*. Teil 1. Exemplarische Analysen des Sprachverhaltens in Mannheim. Berlin: Walter de Gryuter, 1–38.

Kovács, Magdolna. 2004. Australiansuomalaiset kielenvaihdon kynnyksellä. *Virittäjä* 108, 200–223.

Kristiansen, Tore. 2000. Findes Næstved-københavnsk? *Folkmålsstudier* 39, 207–218.

Kristiansen, Tore & Coupland, Nikolas (toim.). 2011. *Standard Languages and Language Standards in a Changing Europe*. Oslo: Novus Press.

Kuparinen, Olli, Mustanoja, Liisa, Peltonen, Jaakko, Santaharju, Jenni & Leino, Unni-Päivä. 2019. Muutosmallit kolmen aikapisteen pitkittäisaineiston valossa. *Sananjalka* 61, 30–56.

Kurki, Tommi. 2005. *Yksilön ja ryhmän kielen reaaliaikainen muuttuminen. Kielenmuutoksen seuraamisesta ja niiden tarkastelussa käytettävistä menetelmistä*. SKST 1036. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kurki, Tommi. 2018. Kielikäsitysten mosaiikki. *Sananjalka* 60, 71–95.
Laakso, Johanna, Sarhimaa, Anneli, Spiliopoulou Åkermark, Sia & Toivanen, Reetta. 2016. *Towards Openly Multilingual Policies and Practices. Assessing Minority Language Maintenance across Europe*. Bristol: Multilingual Matters.
Labov, William. 1966. *The Social Stratification of English in New York City*. Washington (DC): Center for Applied Linguistics.
Labov, William. 1994. *Principles of Linguistic Change. Volume I. Internal Factors*. Oxford: Blackwell.
Lainio, Jarmo. 1989. *Spoken Finnish in Urban Sweden*. Uppsala multiethnic papers 15. Uppsala: Centre for Multiethnic Research.
Lappalainen, Hanna. 2001. Sosiolingvistinen katsaus suomalaisnuorten nykypuhekieleen ja sen tutkimukseen. *Virittäjä* 105, 74–101.
Lappalainen, Hanna. 2004. *Variaatio ja sen funktiot. Erään sosiaalisen verkoston jäsenten kielellisen variaation ja vuorovaikutuksen tarkastelua*. SKST 964. Helsinki: Suomalaisen Kirjallisuuden Seura.
Lappalainen, Hanna. 2006. *Mie* vai *mää*, *sinä* vai *te*? Virkailijoiden kielelliset valinnat itseen ja vastaanottajaan viitattaessa. Julkaisussa: Sorjonen, Marja-Leena & Raevaara, Liisa (toim.) *Arjen asiointia. Keskusteluja Kelan tiskin äärellä*. Tietolipas 210. Helsinki: Suomalaisen Kirjallisuuden Seura, 241–284.
Lappalainen, Hanna. 2011. Förändring, variation och variationens funktioner. En fallstudie av variationen i bruket av personliga pronomen på en finsk landsbygdsort. Julkaisussa: Bockgård, Gustav & Nilsson, Jenny (toim.) *Interaktionell dialektologi*. Uppsala: Institutet för språk och folkminnen, 159–197.
Lappalainen, Hanna, Mustanoja, Liisa & O'Dell, Michael. 2019. Miten ja milloin yksilön kieli muuttuu? Helsinkiläisidiolektien muutos ja muutoksen tutkimuksen menetelmät. *Virittäjä* 123, 550–581.
Latomaa, Sirkku & Nuolijärvi, Pirkko. 2005. The language situation in Finland. Julkaisussa: Kaplan, Robert B. & Baldauf Jr., Richard B. (toim.) *Language Planning and Policy in Europe. Vol. Hungary, Finland and Sweden*. Language planning & policy. Clevedon: Multilingual Matters Ltd, 125–232.
Lehtonen, Heini. 2015. *Tyylitellen. Nuorten kielelliset resurssit ja kielen sosiaalinen indeksisyys monietnisessä Helsingissä*. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-1333-7.
Lehtonen, Heini & Räty, Reetta. 2018. Kielitietoisia käytänteitä monikielisessä koulussa: kokemuksia toimintatutkimuksesta. *Kieli, koulutus ja yhteiskunta* 9:3. Saatavissa: https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-toukokuu-2018/kielitietoisia-kaytanteita-monikielisessa-koulussa-kokemuksia-toimintatutkimuksesta.
Leskinen, Heikki. 1974. Karjalaisen siirtoväen murteen muuttumisesta ja sen tutkimisesta. *Virittäjä* 78, 361–378.
Leskinen, Heikki. 1980. Wie verschwindet ein Dialekt? Beobachtungen zur heutigen Umgangssprache der karelischen Umsiedler. Julkaisussa: Leskinen, Heikki (toim.) *Heutige Wege der finnischen Dialektologie*. Studia Fennica 24. Helsinki: Suomalaisen Kirjallisuuden Seura, 67–91.

Lindgren, Anna-Riitta. 2000. *Helsingin saamelaiset ja oma kieli*. SKST 801. Helsinki: Suomalaisen Kirjallisuuden Seura.
Long, Daniel & Preston, Dennis R. (toim.). 2000. *Handbook of Perceptual Dialectology. Volume 2*. Amsterdam: John Benjamins.
Maegaard, Marie, Gregersen, Frans, Quist, Pia & Jørgensen, J. Normann (toim.). 2009. *Language Attitudes, Standardization and Language Change*. Oslo: Novus.
Mantila, Harri. 2004. Murre ja identiteetti. *Virittäjä* 108, 322–346.
Mielikäinen, Aila (toim.). 1980. *NPMJy, raportti 1*. Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 20. Jyväskylä: Jyväskylän yliopisto.
Mielikäinen, Aila. 1981a. Murre, kielenkäyttäjä ja asenteet. Julkaisussa: Mielikäinen, Aila (toim.) *NPMJy, raportti 2*. Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 24. Jyväskylä: Jyväskylän yliopisto, 91–126.
Mielikäinen, Aila. (toim.). 1981b. *NPMJy, raportti 2*. Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 24. Jyväskylä: Jyväskylän yliopisto.
Mielikäinen, Aila (toim.). 1981c. *NPMJy, raportti 3*. Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 26. Jyväskylä: Jyväskylän yliopisto.
Mielikäinen, Aila. 1982. Nykypuhesuomen alueellista taustaa. *Virittäjä* 86, 277–294.
Mielikäinen, Aila (toim.). 1986. *NPMJy, raportti 4*. Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 32. Jyväskylä: Jyväskylän yliopisto.
Mielikäinen, Aila & Palander, Marjatta. 2014. *Miten suomalaiset puhuvat murteista? Kansanlingvistinen tutkimus metakielestä*. Suomi 203. Helsinki: Suomalaisen Kirjallisuuden Seura.
Milroy, Lesley. 1980. *Language and Social Networks*. Language in Society 2. Oxford: Basil Blackwell.
Milroy, James & Milroy, Lesley. 1997. Varieties and variation. Julkaisussa: Bayley, Robert, Cameron, Richard & Ceil, Lucas (toim.) *The Handbook of Sociolinguistics*. Oxford: Blackwell, 47–64.
Mononen, Kaarina. 2013. *Inkerinsuomalaisten suomen kielen käyttö Pietarissa ja sen lähialueilla*. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-8657-1.
Mustanoja, Liisa. 2011. *Idiolekti ja sen muuttuminen. Reaaliaikatutkimus Tampereen puhekielestä*. Acta Universitatis Tamperensis 1605. Tampere: Tampere University Press.
Nevalainen, Terttu & Raumolin-Brunberg, Helena (toim.). 2003. *Historical Sociolinguistics. Language Change in Tudor and Stuart England*. London: Longman.
Nordberg, Bengt. 1971. Böjningen av neutrala substantiv i Eskilstunaspråket. Utomspråkligt och inomspråkligt betingad variation. *Nysvenska studier* 51, 117–227.
Nordberg, Bengt (toim.). 1994. *The Sociolinguistics of Urbanization. The Case of the Nordic Countries*. Berlin & New York (NY): Walter de Gruyter.
NPMJy = Nykysuomalaisen puhekielen murros. Jyväskylän osatutkimus.
Nuolijärvi, Pirkko. 1986a. *Kieliyhteisön vaihto ja muuttajan identiteetti*. Tietolipas 100. Helsinki: Suomalaisen Kirjallisuuden Seura.
Nuolijärvi, Pirkko. 1986b. *Kolmannen sukupolven kieli. Helsinkiin muuttaneiden suurten ikäluokkien eteläpohjalaisten ja pohjoissavolaisten kielellinen sopeutuminen*. SKST 436. Helsinki: Suomalaisen Kirjallisuuden Seura.

Nuolijärvi, Pirkko. 2000. Sosiolingvistiikka kielentutkimuksen kentässä. Julkaisussa: Sajavaara, Kari & Piirainen-Marsh, Arja (toim.) *Kieli, diskurssi ja yhteisö.* Soveltavan kielentutkimuksen teoriaa ja käytäntöä 2. Soveltavan kielentutkimuksen keskus. Jyväskylä: Jyväskylän yliopisto, 13–37.

Nuolijärvi, Pirkko. 2001. Begreppet stadsmål och den urbana dialektologin. Julkaisussa: Blomqvist, Marianne (toim.) *Våra språk i tid och rum.* Meddelanden från Institutionen för nordiska språk och nordisk litteratur vid Helsingfors universitet utgivna genom Mirja Saari B:21. Helsingfors: Helsingfors universitet, 123–133.

Nuolijärvi, Pirkko & Sorjonen, Marja-Leena. 2005. *Miten kuvata muutosta? Puhutun kielen tutkimuksen lähtökohtia murteenseuruuhankkeen pohjalta*. Helsinki: Kotimaisten kielten tutkimuskeskus.

Nuolijärvi, Pirkko & Tiittula, Liisa. 2000. *Televisiokeskustelun näyttämöllä. Televisioinstitutionaalisuus suomalaisessa ja saksalaisessa keskustelukulttuurissa*. SKST 768. Helsinki: Suomalaisen Kirjallisuuden Seura.

Nuolijärvi, Pirkko & Vaattovaara, Johanna. 2011. De-standardisation in progress in Finnish society? Julkaisussa: Kristiansen, Tore & Coupland, Nikolas (toim.) *Standard Languages and Language Standards in a Changing Europe*. Oslo: Novus Press, 67–74.

Nupponen, Anne-Maria. 2011. *"Savon murre" savolaiskorvin. Kansa murteen havainnoijana*. Joensuu: Itä-Suomen yliopisto.

Omdal, Helge. 1994. *Med språket på flyttefot. Språkvariasjon og språkstrategier blant setesd*øler i Kristiansand. Skrifter utgivna av institutionen för nordiska språk vid Uppsala universitet 35. Uppsala: Uppsala universitet.

Palander, Marjatta. 2005. *Lapsuudesta keski-ikään. Seuruututkimus itäsavolaisen yksilömurteen kehityksestä*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Palander, Marjatta. 2011. *Itä- ja eteläsuomalaisten murrekäsitykset.* Suomi 200. Helsinki: Suomalaisen Kirjallisuuden Seura.

Paunonen, Heikki. 1995 [1982]. *Suomen kieli Helsingissä. Huomioita Helsingin puhekielen historiallisesta taustasta ja nykyvariaatiosta*. Helsinki: Helsingin yliopiston suomen kielen laitos.

Paunonen, Heikki. 1995. Puhesuomen muuttuva omistusmuotojärjestelmä. *Virittäjä* 99, 501–531.

Paunonen, Heikki. 1996. Language change in apparent time and in real time. Possessive constructions in Helsinki colloquial Finnish. Julkaisussa: Thelander, Mats, Elmevik, Lennart, Gunnarsson, Britt-Louise, Melander, Björn (toim.) *Samspel och variation. Språkliga studier tillägnade Bengt Nordberg på 60-årsdagen*. Uppsala, 375–386.

Paunonen, Heikki. 2005. Helsinkiläisiä puhujaprofiileja. *Virittäjä* 109, 162–200.

Paunonen, Heikki. 2018. Tornion murteesta meänkieleen. Reaaliaikainen tutkimus Ruotsin Ylitorniolta 1966–1992. Julkaisussa: Brunni, Sisko, Kunnas, Niina, Palviainen, Santeri & Sivonen, Jari (toim.) *Kuinka mahottomasti nää tekkiit*. Juhlakirja Harri Mantilan 60-vuotispäivän kunniaksi. Studia humaniora ouluensia 16. Oulu: Oulun yliopisto, 15–161.

Paunonen, Heikki & Paunonen, Marjatta. 2000. *Tsennaaks Stadii, bonjaaks slangii. Stadin slangin suursanakirja*. Helsinki: WSOY.

Pedersen, Inge Lise. 1986. Bondsk och ligefrem eller moderne og byagtig. En undersøgelse af holdninger til dialekt og dialekttalende på Fyn. *Danske Folkemål* 28, 11–39.

Peterson, Elizabeth & Vaattovaara, Johanna. 2014. *Kiitos* and *pliis*. The relationship of native and borrowed politeness markers in Finnish. *Journal of Politeness Research* 10, 247–269.

Piippo, Irina. 2012. *Viewing Norms Dialogically. An Action-Oriented Approach to Sociolinguistic Metatheory*. Helsinki: University of Helsinki.

Piippo, Irina, Vaattovaara, Johanna & Voutilainen, Eero (toim.). 2016. *Kielen taju. Vuorovaikutus, asenteet ja ideologiat*. Helsinki: Art House.

Preston, Dennis R. 1989. *Perceptual Dialectology. Nonlinguists' Views of Areal Linguistics*. Dordrecht: Foris Publications.

Preston, Dennis R. (toim.). 1999. *Handbook of Perceptual Dialectology. Volume 1*. Amsterdam: John Benjamins.

Priiki, Katri. 2017. *Hän, se, tää* vai *toi? Vuorovaikutussosiolingvistinen tutkimus henkilöviittauksista Kaakkois-Satakunnan nykypuhekielessä*. Turun yliopiston julkaisuja. Sarja C, osa 432. Scripta Lingua Fennica Edita. Turku: Turun yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-29-6719-3.

Priiki, Katri, Mononen, Kaarina & Lappalainen, Hanna. 2018. Monitieteinen tutkimushanke nykysuomalaisten kielellisistä elämäkerroista. *Sananjalka* 60, 260–267.

Rampton, Ben. 1995. *Crossing. Language and Ethnicity among Adolescents*. London: Longman.

Rampton, Ben. 2006. *Language in Late Modernity. Interaction in an Urban School*. Cambridge: Cambridge University Press.

Rouhikoski, Anu. 2015. *Laita, laitatko* vai *laitat?* Kolmen direktiivirakenteen variaatio asiakaspalvelutilanteessa. *Virittäjä* 119, 189–22.

Sarhimaa, Anneli. 2017. *Vaietut ja vaiennetut. Karjalankieliset karjalaiset Suomessa*. Tietolipas 256. Helsinki: Suomalaisen Kirjallisuuden Seura.

Siitonen, Kirsti. 1981. Dialektologian ja sosiolingvistiikan suhteesta. Julkaisussa: Suojanen, M. K. (toim.) *Kirjoituksia puhekielestä*. Turun puhekielen projektin julkaisuja 1. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 14. Turku: Turun yliopisto, 43–58.

Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language and Communication* 23, 193–229.

SKST = Suomalaisen Kirjallisuuden Seuran Toimituksia.

Sorjonen, Marja-Leena, Rouhikoski, Anu & Lehtonen, Heini (toim.). 2015. *Helsingissä puhuttavat suomet. Kielen indeksisyys ja sosiaaliset identiteetit*. SKST 1310. Helsinki: Suomalaisen Kirjallisuuden Seura.

Sundgren, Eva. 2002. Återbesök i Eskilstuna. En undersökning av morfologisk variation och förändring i nutida talspråk. Uppsala: Uppsala universitet.

Suojanen, M. K. (toim.). 1981. *Kirjoituksia puhekielestä*. Turun puhekielen projektin julkaisuja 1. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 14. Turku: Turun yliopisto.

Suojanen, M. K. 1985. *Mitä Turussa puhutaan? Raportti Turun puhekielen tutkimuksesta*. Turun puhekielen projektin julkaisuja 3. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 23. Turku: Turun yliopisto.

Suojanen, M. K. & Suojanen, Päivikki (toim.). 1982. *Sosiolingvistiikan näkymiä*. Helsinki: Gaudeamus.

Tagliamonte, Sali A. 2006. *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Tandefelt, Marika. 1988. *Mellan två språk. En fallstudie om språkbevarande och språkbyte i Finland*. Acta Universitatis Upsaliensis. Studia Multiethnica Upsaliensia 3. Uppsala: Uppsala universitet.

Thelander, Mats. 1979a. *Språkliga variationsmodeller tillämpade på nutida Burträsktal. Del 1. Materialbeskrivning och analys på språkdragsnivå*. Studia Philologiae Scandinaviae Upsaliensia 14: 1. Uppsala: Almqvist & Wiksell.

Thelander, Mats. 1979b. *Språkliga variationsmodeller tillämpade på nutida Burträsktal. Del 2. Teoretiska utgångspunkter och analys på språkartsnivå*. Studia Philologiae Scandinaviae Upsaliensia 14: 2. Uppsala: Almqvist & Wiksell.

Trudgill, Peter. 1974. *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Vaattovaara, Johanna. 2009. *Meän tapa puhua. Tornionlaakso pellolaisnuorten subjektiivisena paikkana ja murrealueena*. SKST 1224. Helsinki: Suomalaisen Kirjallisuuden Seura.

Vaattovaara, Johanna. 2014. On the dynamics of non-linguists' dialect perceptions – the perceived spatiality of /s/ variation in Finnish. Julkaisussa: Reif, Monika, Robinson, Justyna A. & Pütz, Martin (toim.) *Variation in Language and Language Use. Linguistic, Socio-Cultural and Cognitive Perspectives*. Duisburg Papers on Research in Language and Culture. Frankfurt am Main: Peter Lang, 134–163.

Vaattovaara, Johanna. 2016. Murtuvat murteet ja murreasenteet. Julkaisussa: Piippo, Irina, Vaattovaara, Johanna & Voutilainen, Eero (toim.) *Kielen taju. Vuorovaikutus, asenteet ja ideologiat*. Helsinki: Art House, 83–105.

Vaattovaara, Johanna & Halonen, Mia. 2015. Missä on ässä? "Stadilaisen s:n" helsinkiläisyydestä. Julkaisussa: Sorjonen, Marja-Leena, Rouhikoski, Anu & Lehtonen, Heini (toim.) *Helsingissä puhuttavat suomet. Kielen indeksisyys ja sosiaaliset identiteetit*. SKST 1310. Helsinki: Suomalaisen Kirjallisuuden Seura, 40–83.

Vaattovaara, Johanna, Kunnas, Niina & Saviniemi, Maija. 2018. Stadi imitoituna. Julkaisussa: Brunni, Sisko, Kunnas, Niina, Palviainen, Santeri & Sivonen, Jari (toim.) *Kuinka mahottomasti nää tekkiit*. Juhlakirja Harri Mantilan 60-vuotispäivän kunniaksi. Studia humaniora ouluensia 16. Oulu: Oulun yliopisto, 287–323.

Vaattovaara, Johanna & Poutiainen, Saila. 2017. Kieli-ideologiat kierrätyksessä: tapaustutkimus "lässytys"-uutisesta ja sen kommenttiketjusta. *AFinLAn vuosikirja* 75, 133–160.

Wagner, Suzanne Evans & Buchstaller, Isabelle (toim.). 2018. *Panel Studies of Variation and Change*. London: Routledge.

# Luku 22 Keskustelunanalyysi ja vuorovaikutuslingvistiikka

*Katariina Harjunpää*
https://orcid.org/0000-0002-4586-1563

*Jarkko Niemi*
https://orcid.org/0000-0001-6747-7821

*Marja-Leena Sorjonen*
https://orcid.org/0000-0002-8225-6805

**Mikä?**

Keskustelunanalyysissa ja vuorovaikutuslingvistiikassa tutkitaan kieltä ja vuorovaikutusta erilaisissa sosiaalisissa tilanteissa ja yhteiskunnan eri alueilla. Kiinnostus kohdistuu siihen, miten ihmisten välinen vuorovaikutus, toiminta ja yhteisymmärrys rakentuvat: kieltä tutkitaan sosiaalisen vuorovaikutuksen ja toiminnan jäsentymisen kautta. Tutkimus voi koskea esimerkiksi vaihtoehtoisia kielellisiä tapoja muotoilla tietty toiminto tai toimintajakso, kielen suhdetta muihin vuorovaikutuksen resursseihin, kielellisiä kategorioita vuorovaikutuksen näkökulmasta tai jonkin vuorovaikutustilanteen kokonaisuuden jäsentymistä.

**Katso myös:**

Johdannon luku 4. Kielitieteen etiikka – hyvä tutkimus murroksessa

Luku 4 Kokeellinen fonetiikka

Luku 5 Lingvistinen etnografia toimintayhteisöissä

Luku 6 Diskurssianalyysi

Luku 20 Tekstintutkimus

Luku 21 Sosiolingvistiikka

# 1. Johdanto

Keskustelunanalyyttisen ja vuorovaikutuslingvistisen tutkimuksen kohteena on sosiaalinen vuorovaikutus ja kieli erilaisissa tilanteissa ja yhteiskunnan alueilla – esimerkiksi aamiaisella kotona, isoäidin syntymäpäivillä, sairaalan lääkärinkierrolla, työvoimatoimistossa, luokkahuoneessa, televisiokeskustelussa, internet- tai lautapeliä pelattaessa. Keskustelunanalyyttisia ja vuorovaikutuslingvistisiä menetelmiä voi soveltaa myös kirjoitetun vuorovaikutuksen tutkimukseen (esim. kaunokirjallinen dialogi, WhatsApp- ja muu pikaviestiminen). Näillä kahdella tutkimussuuntauksen nimellä viitataan hieman erilaisiin tutkimusnäkökulmiin ja -perinteisiin. Keskustelunanalyysi on lähtöisin sosiologiasta, kun taas vuorovaikutuslingvistiikka on myöhemmin kehittynyt kielentutkimuksen suuntaus, jossa keskustelunanalyyttiset menetelmät ovat keskeisiä. Tilanteen mukaan samakin tutkija voi esitellä itsensä joko keskustelunanalyytikkona, vuorovaikutuslingvistinä tai lingvistisesti suuntautuneena keskustelunanalyytikkona. Keskustelunanalyyttista tutkimusta tehdään ja hyödynnetään nykyisin monilla tieteenaloilla, kielentutkimuksen, sosiologian ja muiden sosiaalitieteiden lisäksi esimerkiksi puhetieteissä, kasvatustieteessä, antropologiassa ja neurotieteissä tutkittaessa vuorovaikutuksen jäsentymistä.

Keskustelunanalyytikkojen ja vuorovaikutuslingvistien kiinnostuksen kohteena on se, millaisten käytänteiden ja toimintojen avulla sosiaalista vuorovaikutusta rakennetaan ja jäsennetään. Vuorovaikutuksen yksityiskohtia tarkastelemalla pyritään ymmärtämään sellaisia osallistujien kompetensseja, jotka muodostavat perustan jokapäiväisille sosiaalisille toiminnoille (Heritage 1996 [1984], 236–239). Analyysin avulla päästään kiinni siihen, mikä tekee sosiaalisen vuorovaikutuksen ja osallistujien välisen yhteisymmärryksen mahdolliseksi. Lähtökohtana on, että tutkitaan autenttisia vuorovaikutustilanteita, siis sellaisia, joita ei ole järjestetty erikseen tutkimusta varten. Aineistona voivat olla puhutuista tai viittomakielisistä keskusteluista tehdyt ääni- ja videotallenteet tai teknologiavälitteisten puhuttujen ja kirjoitettujen keskustelujen tallenteet. Myös fiktiivisiä keskusteluja (esim. kaunokirjalliset dialogit, näytelmät, televisiosarjat) voi tutkia keskustelunanalyysin ja vuorovaikutuslingvistiikan menetelmillä. Tutkitaan sitä, miten keskustelijat osoittavat toisilleen erilaisin kielellisin ja kehollisin keinoin suuntautumista siihen, mitä ollaan tekemässä ja millainen on heidän välinen suhteensa esimerkiksi tiedollisesti tai affektisesti tapahtumahetkellä.

Tutkimuksessa voidaan lähteä liikkeelle eri tavoin riippuen kiinnostuksen kohteesta. Tutkija saattaa olla kiinnostunut eri tavoista, joilla puhujat muotoilevat jonkin toiminnon, vaikka pyynnön tai pääasiaan palaamisen. Vaihtoehtoisesti kiinnostus voi kohdistua johonkin kielenainekseen, johon tutkija on kiinnittänyt huomiota (esim. kysymyksen aloittaminen *and*-konjunktiolla, Heritage & Sorjonen 1994; läsnäolijaan viittaaminen demonstratiivilla *tuo*, Seppänen 1998; johonkin tarkoitteeseen viittaaminen demonstratiivilla *tämä*, Etelämäki 2006; -*s*-liitepartikkeli kysymyksissä, Raevaara 2004). Tarkasteltavana voi olla myös jokin kielellinen kategoria ja pohdittavana se, miten keskusteluaineiston analyysi suhteutuu kielentutkimuksen aiempaan kuvaukseen kategoriasta (esim. alisteisista lauseista, Laury & Suzuki 2011; menneen ajan aikamuodoista, Seppänen 1997a; subjektin merkinnästä, Helasvuo 2014; *yksikön* käsitteestä kielentutkimuksessa, Ono, Laury & Suzuki 2019; responsiivisten toimintojen yksiköistä, Helasvuo, Endo & Kärkkäinen 2018). Käytännössä tutkija vaihtelee analyysin fokusta kielenainesten ja toiminnan välillä tutkimuksen kuluessa. Metodi tarjoaa

vuorovaikutuksen tutkimiselle vankan empiirisen pohjan, jota voidaan haluttaessa täydentää muiden alojen näkemyksillä. Keskustelunanalyyttisesti päästään käsiksi vaikkapa sellaisiin ilmiöihin kuin tietäminen, asennoituminen, päätöksenteko, yhteisymmärryksen syntyminen ja erimielisyys, oppiminen, asiantuntijuus, sosiaalinen kategorisointi ja yhdessä liikkuminen – vuorovaikutuksen perusjäsennyksillä on osansa näissä kaikissa.

Etenemme tässä artikkelissa seuraavasti: Selvitämme ensin keskustelunanalyysin ja vuorovaikutuslingvistiikan syntyä, mukaan lukien suomalainen tutkimus (luku 2), ja kuvaamme tutkimusalojen teoreettis-metodisia lähtökohtia (luku 3). Sen jälkeen esittelemme ja havainnollistamme analyysimenetelmiä (luku 4). Päätämme artikkelin hahmottelemalla joitakin alan ajankohtaisia kysymyksiä (luku 5).

## 2. Keskustelunanalyysin ja vuorovaikutuslingvistiikan historiaa ja taustaa

### 2.1. Etnometodologiasta keskustelunanalyysiin

Keskustelunanalyysi syntyi 1960-luvun lopulla amerikkalaisessa sosiologiassa Harold Garfinkelin kehittämän etnometodologian pohjalta, ja taustalla oli myös sosiologi Erving Goffmanin ajattelu. Garfinkel oli tyytymätön aikansa sosiologian tapaan nähdä sosiaalisen toiminnan ja yhteiskunnan järjestyneisyys tulokseksi yksilöiden sisäistämistä, ylhäältä annetuista normeista. Hän esitti, että ihmisten välisessä toiminnassa ja yhteisymmärryksen saavuttamisessa keskeisiä ovat toimijoiden jakamat arkipäättelyn metodit, joiden avulla he tuottavat ymmärryksen toistensa toiminnasta vuorovaikutuksen kuluessa (ks. Garfinkel 1967; Heritage 1996 [1984]). Samoihin aikoihin sosiologi Erving Goffman esitti näkemyksiään sosiaalisten tilanteiden rakentumisesta ja yksilön suhteesta muihin. Hänen mukaansa näiden sosiaalisten ilmiöiden jäsentyminen

kasvokkaisessa vuorovaikutuksessa (engl. *interaction order*) on oma tutkimuskohteensa (Goffman 1967, 2–3; 1983; 2012). Keskustelunanalyysin perustaja Harvey Sacks oli sekä Goffmanin että etenkin Garfinkelin oppilas (ks. esim. Schegloff 1992; Koskela & Piirainen-Marsh 2002; Maynard 2012).[1]

Sacks vei Garfinkelin ajattelua sosiaalisten toimijoiden jakamista arkipäättelyn metodeista uudelle alueelle, sosiaalisen vuorovaikutuksen empiiriseen analyysiin. Kimmoke tähän olivat kriisipuhelimeen soitetut puhelut, joihin Sacks tutustui työskennellessään itsemurhien ehkäisyyn keskittyneessä tutkimuslaitoksessa Los Angelesissa. Hän kiinnostui niistä systemaattisista ja jaetuista kulttuurisista menetelmistä, metodeista, joilla keskustelijat jäsentävät vuorovaikutusta ja joilla he osoittavat suuntautumistaan toistensa toimintoihin ja niiden relevanssiin vuorovaikutuksen kuluessa – kieli sinänsä ei ollut Sacksin ensisijainen mielenkiinnon kohde.

Sacks kehitti keskustelunanalyyttista ohjelmaa Kalifornian yliopistossa 1964–1972 luennoissaan, jotka julkaistiin postuumisti Gail Jeffersonin toimittamina (Sacks 1992). 1970-luvulla julkaistiin keskustelunanalyysin klassiset artikkelit vuorovaikutuspuheen jäsentymisestä, sen infrastruktuurin muodostavista perusjäsennyksistä. Artikkelit käsittelivät keskustelun jäsentymistä toimintakokonaisuuksiksi, puheen vuorottelua ja vuorovaikutuksen ongelmien käsittelyä (Schegloff & Sacks 1973; Sacks, Schegloff & Jefferson 2016 [1974]; Schegloff, Jefferson & Sacks 1977).[2]

**Keskustelun perusjäsennykset**

Alkuvaiheistaan lähtien keskustelunanalyysi on pyrkinyt osoittamaan, millaisten perustavien jäsennyskeinojen varaan puhuttu vuorovaikutus rakentuu. *Vuorottelujäsennys* kuvaa mekanismeja, joilla osallistujat järjestävät puheenvuorojen ottamisen ja vaihtumisen (Sacks, Schegloff & Jefferson 2016 [1974]). *Sekvenssijäsennys* kattaa toimintajaksojen perusrakenteet, joilla puhujien vuorot suhteutuvat toisiinsa ja muodostavat koherentteja sosiaalisen toiminnan kokonaisuuksia (ks. Schegloff 2007; Haakana 2008; Raevaara 2016). *Korjausjäsennyksen* avulla osallistujat kä-

sittelevät puhumisen, kuulemisen ja ymmärtämisen ongelmia (ks. Sorjonen 1997; Hayashi, Raymond & Sidnell 2013; Haakana, Kurhila, Lilja & Savijärvi 2016). Vuorovaikutuksen jäsentymisen perusilmiöihin kuuluvat myös vastaanottajan huomioiva muotoilu (engl. *recipient design*, Sacks, Schegloff & Jefferson 2016 [1974]: 50) eli se, miten puhuja muotoilee vuoronsa tilanteeseen ja tietylle vastaanottajalle sopivaksi, sekä osallistumisen ja yhteistoiminnan kehollinen koordinointi (Seppänen 1997c; C. Goodwin & M. H. Goodwin 2004; C. Goodwin 2017).

Sacksin menehdyttyä auto-onnettomuudessa vuonna 1975 keskustelunanalyysin kehittämistä jatkoivat keskeisesti Emanuel Schegloff ja Gail Jefferson. Keskustelunanalyysin tuntemusta ja asemaa kielentutkimuksessa vahvisti Stephen C. Levinsonin (1983) oppikirja *Pragmatics*, jonka keskustelunanalyysia käsittelevässä laajassa luvussa Levinson tekee eron kirjoittamisajan diskurssianalyysin[3] ja keskustelunanalyysin välille ja osoittaa keskustelunanalyysin kontribuoivan tärkeällä tavalla ymmärrykseen lausuman merkityksestä (esim. epäsuorat puheaktit, Gricen maksiimit, presuppositio, deiksis; ks. Levinson 1983, luku 8). Sittemmin keskustelunanalyysista on tullut keskeinen vuorovaikutuksen empiirisen tutkimuksen suuntaus useilla tieteenaloilla (ks. Sidnell & Stivers 2012).

Suomeen keskustelunanalyysin toivat 1980-luvulla kielentutkijat, Auli Hakulinen ja hänen tutkimusryhmänsä.[4] Hakulinen oli kiinnostunut arkipuheen syntaksista, ja sen tutkimiseksi haettiin menetelmiä, joilla tutkia puhuttua vuorovaikutusta. Konkreettinen metodinen ponnin oli Ruotsissa 1987 järjestetty pohjoismainen kesäkoulu, jonka opettajiin kuuluivat Schegloff, John Heritage ja Adam Kendon. Sen tuloksena alkoi suhteiden solmiminen keskustelunanalyysin keskeisiin tutkijoihin, joista monet ovat vierailleet kouluttajina Suomessa (ensimmäisinä Gail Jefferson[5], Paul Drew sekä Charles ja Marjorie Harness Goodwin). Vuonna 1989 ilmestyi tutkimusryhmän kokoomateos *Suomalaisen keskustelun keinoja* (Hakulinen 1989), joka sisältää Hakulisen kirjoittaman laajan jakson keskustelunanalyysin soveltuvuudesta keskustelupuheen tutkimusmenetelmäksi sekä moniulotteisen kirjoituksen keskustelun

tutkimuksen tavoitteista ja puheen syntaksin ja perinteisen kieliopin tutkimuksen suhteista (luvut 1–2).[6] Teos sisältää kirjoituksia modaaliverbeistä, henkilöviittauksista, puheenvuoroa aloittavista aineksista ja myös menetelmällisistä kysymyksistä. 1980-luvun lopulla keskustelunanalyyttista tutkimusta tekivät ja hyödynsivät fennistien lisäksi myös logopedit, sosiolingvistit ja suomenruotsin tutkijat. 1990-luvun alkupuolella sai alkunsa ensimmäinen sosiaalitieteilijöiden ja kielentutkijoiden yhteishanke vuorovaikutuksesta lääkärin vastaanotolla (ks. Hakulinen & Sorjonen 1993). 2000-luvun alkuun tultaessa keskustelunanalyyttista tutkimusta tekeviä tutkijoita oli lähes kaikissa maamme yliopistoissa ja useilla tieteenaloilla.

## 2.2. Keskustelunanalyysista vuorovaikutuslingvistiikkaan

Termin *vuorovaikutuslingvistiikka* (engl. *interactional linguistics*) lanseerasivat 2000-luvun alussa Saksassa toimineet kielentutkijat kattotermiksi tutkimukselle, joka selvitti keskustelupuheen fonologiaa, kieliopin ja vuorovaikutuksen suhdetta, kielen kontekstualisointia ja keskustelupuheen prosodiaa (Couper-Kuhlen & Selting 2001, 1–22). Suuntauksen lähtölaukauksena toimi kaksi perättäistä Eurokonferenssia (EuroConference; Spa 2000, Linguistic Organization of Conversational Activities ja Helsinki 2002, Linguistic Structures and their Deployment in the Organisation of Conversation).[7]

Vuorovaikutuslingvistiikan synnyn taustalla oli kolme toisistaan erillistä tutkijaverkostoa: Isossa-Britanniassa keskustelupuheen fonetiikkaan keskittyneet tutkijat, Saksassa keskustelupuheen syntaksiin ja prosodiaan keskittyneet tutkijat sekä Yhdysvalloissa (diskurssi)funktionaalisen syntaksin sekä kieliopin ja vuorovaikutuksen tutkijat (Couper-Kuhlen & Selting 2017, 4–14; myös Clift 2016, 26–27). Kieliopin ja vuorovaikutuksen tutkijoiden moninainen tausta näkyi kokoomateoksessa *Interaction and grammar* (1996). Sen toimittajat Elinor Ochs, Emanuel Schegloff ja Sandra Thompson toimivat kolmella tieteenalalla, lingvistisessä antropologiassa, sosiologiassa ja funktionaalisessa kielitieteessä. He toteavat teoksen johdannossa, että kielioppia lähestytään

teoksessa osana laajempaa vuorovaikutusresurssien joukkoa. Kielioppi nähdään yhtenä resurssina käytänteissä, jotka muodostavat perustan sosiaaliselle elämälle (esim. toiminnan, tiedon ja tunteiden ilmaisun jäsentymiselle) ja erityisesti sille, miten kieli on mukana arjen vuorovaikutuksessa (Schegloff, Ochs & Thompson 1996).

Couper-Kuhlen ja Selting (2017, 16) kiteyttävät kolme vuorovaikutuslingvistiikan yleistä tavoitetta. Tavoitteena on 1) tarjota funktionaalinen kuvaus kielellisistä rakenteista resursseina, joilla hoidetaan toistuvia tehtäviä sosiaalisessa vuorovaikutuksessa, 2) analysoida ja verrata eri kielten käytänteitä, jotta saadaan selville, kuinka kielen rakenne vaikuttaa sosiaalisen vuorovaikutuksen yksityiskohtiin ja kuinka vuorovaikutuksen vaatimukset muokkaavat kielen rakennetta ja käyttöä ja 3) rakentaa sellaista teoriaa, joka selittää, miten kieli on jäsentynyt ja miten sitä käytetään sosiaalisessa vuorovaikutuksessa.

Vuorovaikutuslingvistisen tutkimuksen metodinen kirjo on laaja esimerkiksi sen suhteen, mikä osuus on vuorovaikutuksen ja kielen toimintoluonnetta korostavilla keskustelunanalyyttisilla metodeilla, mikä kielen kategorioihin keskittyvällä perinteisemmin lingvistisellä tutkimuksella. Vastaavasti laadullisten ja määrällisten menetelmien käyttäminen vaihtelee. Vuorovaikutuslingvistiikan synnyn myötä suomalaiset tutkijat ovat asemoineet tutkimuksensa metodisesti joko vuorovaikutuslingvistiseksi tai keskustelunanalyyttiseksi tai molempiin suuntauksiin kuuluvaksi (esim. Etelämäki 2006, 12; Korpela 2007, 15; Helasvuo 2008; Koivisto 2011, 13–16; Niemi 2015, 7–8; Heinonen 2017, 11–12).

## 2.3. Keskustelunanalyysista multimodaaliseen keskustelunanalyysiin

Mitä enemmän videotallenteiden käyttö on yleistynyt keskustelunanalyyttisessa tutkimuksessa, sitä enemmän tutkijat ovat alkaneet suuntautua vuorovaikutuksen kokonaisvaltaisempaan tarkasteluun ottamalla huomioon vuorovaikutuksen resurssit entistä laajemmin: esimerkiksi puheen foneettis-prosodiset piirteet, erilaiset kehollisen sosiaalisen toiminnan muodot sekä fyysisen tilan ja sen materiaalisen luonteen.

Multimodaalisen keskustelunanalyysin uranuurtajia ovat olleet amerikkalaiset antropologit Charles Goodwin (1979; 1981; 2017) ja Marjorie Harness Goodwin (1980; M. H. Goodwin & Cekaite 2018) sekä brittisosiologi Christian Heath (1986), jotka aloittivat vuorovaikutustilanteiden kuvaamisen tutkimusaineistoksi. Charles Goodwin on rakentanut systemaattisesti holistisempaa tutkimusta alkaen katseen, eleiden, ruumiinasennon ja tilankäytön tutkimuksesta ulottuen aina siihen, miten materiaalinen ympäristö esiintyy vuorovaikutuksessa. Hänen varhainen artikkelinsa (1979) lauseen rakentamisesta suhteessa katseella osoitettuihin osallistumiskehikon muutoksiin on ollut vaikutusvaltainen myös vuorovaikutusta kielitieteellisestä näkökulmasta käsittelevässä tutkimuksessa. Erilaisen kehollisen (engl. *embodied*) toiminnan tutkimus on laajentunut 1990-luvulta alkaen niin, että 2000-luvun tilannetta on kuvattu tutkimuksen multimodaalisen käänteen ajaksi keskustelunanalyysissa (Nevile 2015; Mondada 2016).

Suomalaiseen keskustelunanalyyttiseen ja vuorovaikutuslingvistiseen tutkimukseen kehollisen toiminnan ja myös erilaisten materiaalien käytön tarkastelun toivat jo varhain logopedit, jotka analysoivat keskusteluja, joissa oli afaattinen puhuja (Klippi 1996; Laakso 1997). Viime aikoina on tutkittu esimerkiksi opettajajohtoisen luokkahuonevuorovaikutuksen multimodaalisuutta (Kääntä 2015) ja oppilaiden välisiä kosketuksia (Karvonen, Heinonen & Tainio 2018), kielen ja kehollisen toiminnan yhdistelmiä tarjouksen muotoilussa (Kärkkäinen & Keisanen 2012), monitoimintaa (engl. *multiactivity*, Haddington, Keisanen, Mondada & Nevile 2014) sekä kirjoittamista erilaisissa vuorovaikutustilanteissa (Mondada & Svinhufvud 2016). On tutkittu myös vuorovaikutusta erilaisissa ympäristöissä liikuttaessa, kuten autolla ajaessa ja metsässä (Haddington & Rauniomaa 2011; Keisanen, Rauniomaa & Siitonen 2017; Siitonen, Rauniomaa & Keisanen 2019). Multimodaalista tutkimusta erilaisista kauppa-asioinneista (kioskilla, leipomossa, juustokaupassa) on tehty ja tekeillä niin Suomessa kuin monissa muissa Euroopan maissa (esim. Harjunpää, Mondada & Svinhufvud 2018; Lappalainen & Raevaara 2009; Sorjonen & Raevaara 2014; Svinhufvud 2018).

Kieliopin tutkimuksen alueella Eeva-Leena Seppänen (1998) näytti katseen analyysin tuella, että läsnä olevaan henkilöön viittaavat

demonstratiivipronominit *tämä* ja *tuo* ovat keino osoittaa läsnäolijoiden puhujastatuksia eikä niinkään heidän spatiaalisia suhteitaan. Viimeaikaisessa tutkimuksessa taas on muun muassa käynyt ilmi, että lauseen ja lausekkeen vaihtelu keinona pyytää myyjältä tuotetta kioskilla on ikonisessa suhteessa siihen, kuinka kaukana asiakas on suhteellisesti tiskistä ja kuinka kauan häneltä kestää saapua tiskille – kielellistä keinoa muokataan joustavasti meneillään olevaan keholliseen ja sekventiaaliseen toimintaan (Sorjonen & Raevaara 2014). Chiara Monzoni ja Ritva Laury (2015) puolestaan osoittavat, kuinka referenttiä, jota ei ole aiemmin mainittu, arvioidaan ilman mainintaa luomalla pääsy siihen multimodaalisesti. Parhaillaan on käynnissä keskustelu siitä, voidaanko kehollista toimintaa pitää osana kielioppia tietyissä tapauksissa (Keevallik 2018; Couper-Kuhlen 2018; Streeck 2018).

Vuorovaikutuspuheen foneettinen tutkimus keskustelunanalyyttisesta näkökulmasta alkoi Isossa-Britanniassa 1980-luvun alkupuolella, jolloin John Local ja Peter French tutkivat muun muassa vuorosta kilpailun fonetiikkaa (French & Local 1983), ja siitä alkaen vuorovaikutuksen foneettinen tutkimus on jatkunut vahvana (esim. Ogden 2013). Vuorovaikutuksen prosodinen tutkimus sen sijaan sai vahvemmin alkunsa Saksassa (Couper-Kuhlen & Selting 1996; Couper-Kuhlen & Ford 2004). Suomen kielen fonetiikkaa ja prosodiaa on selvittänyt muun muassa Richard Ogden (2001), joka on tutkinut narinan ja glottaaliklusiilin yhteyttä puheenvuoron vaihtamiseen ja sen pitämiseen. Ogden, Auli Hakulinen ja Liisa Tainio (2004) ovat analysoineet prosodista tyylittelyä, jota käytetään vähättelemään sanotun uutisarvoa, ja Ogden ja Sara Routarinne (2005) nousevan intonaation tehtäviä suomessa.

## 2.4. Kielikäsitys

Keskustelunanalyysi ei ole varsinaisesti kieltä koskeva teoria, vuorovaikutuslingvistiikka sen sijaan keskittyy nimenomaisesti kieleen. Keskustelunanalyysin pääasiallinen kiinnostuksen kohde on sosiaalisen toiminnan jäsentyminen ajallisessa ja sekventiaalisessa kontekstissaan. Suuntaus sisältää kuitenkin kielentutkimukselle relevantteja oletuksia.

Yksi niistä on Garfinkelin (1967) esittämä ajatus kaikkien kielellisten ilmausten indeksisyydestä. Se laajentaa perinteistä filosofista ja lingvististä näkökulmaa indeksisyyteen: indeksisyys koskee kaikkea vuorovaikutuspuhetta ja kielellistä toimintaa, niin että kaikkien ilmausten tulkitseminen vaatii kontekstin huomioon ottamista (Garfinkel 1967). Keskustelunanalyyttisessa tutkimuksessa tämä liittyy siihen, että lausuman ja vuoron tulkinnassa olennaista on sekä vuoron muotoilu että sen käyttökonteksti, kuten kielellisen tuotoksen paikka meneillään olevassa vuorossa ja laajemmassa toiminnassa. Usein annettu esimerkki on kysymyslause ja sen tehtävät vuorovaikutuksessa. Esimerkiksi englannin kysymyslause *why don't you lay down and take a nap* voi käyttökontekstissaan toimia kysymyksen sijasta pikemminkin ehdotuksena lopettaa puhelu (Schegloff & Sacks 1973; Thompson & Couper-Kuhlen tulossa). Kysymyslauseen *otatko kahvia* monitehtäväisyys käy puolestaan ilmi vastauksesta *kyllä kiitos*, jolla puhuja vastaa niin lausuman muotoon vaihtoehtokysymyksenä (*kyllä*) kuin sen funktioon tarjouksena (ks. Schegloff 2007, 75–76; myös Schegloff 1984).

Vuorovaikutusta tarkastellaan dynaamisesti ajassa muotoutuvana toimintana, jossa kielelliset rakenteet näyttäytyvät sekventiaaliseen kontekstiinsa sovitettuina käytänteinä. Tutkimuksessa on tässä yhteydessä käytetty käsitettä *paikkasensitiivinen kielioppi* (engl. *positionally sensitive grammar*, Schegloff 1996). Käsitteellä viitataan siihen, miten kielellisten rakenteiden käyttö ja tulkinta kytkeytyvät vuorovaikutuksen ajalliseen ja sekventiaaliseen etenemiseen. Esimerkiksi partikkelin *kyllä* funktio varioi sen mukaan, onko partikkeli lausuman alussa, keskivaiheilla vai lopussa (Hakulinen 2001a). Englannissa puolestaan tarkennusta hakevan kysymyksen (kuten *When's the next elections*) jälkeen lausekevastaus (*Two thousand four*) on tunnusmerkitön eli ongelmattomana käsitelty ja frekvenssiltään yleinen, kun taas kerrontaa hakevan kysymyksen (*How you guys know each other*) jälkeistä lausekevastausta (*Pam*, nimeten yhteisen tuttavan) käsitellään ongelmallisena (Thompson, Fox & Couper-Kuhlen 2015, 23–25, 42–44). Sama vastaustyyppi toimii siis eri tavoin riippuen kysymyksen funktiosta (kysymyksistä ja vastauksista suomessa ks. Sorjonen 2001a; 2001b; Hakulinen 2001b; Vepsäläinen 2019). Monet keskustelunanalyyttiset tutkimukset ovatkin tarkastelleet kielellisten

rakenteiden käyttöä tietyssä sekventiaalisessa asemassa, kuten Niemi (2010) tutkimuksessaan verbialkuisen lauserakenteen käytöstä kahden puhujan välisessä myönnyttelevässä toimintajaksossa ja tutkimuksessaan (2013) neljän muodoltaan läheisen rakenteen käytöstä vastaamassa toteamukseen tai vaihtoehtokysymykseen (myös Couper-Kuhlen & Thompson 2000; muunlaisista tapauksista ks. esim. Curl & Drew 2008; Hakulinen & Sorjonen 2009; Sorjonen & Hakulinen 2009; Koivisto 2017).

Kielellisten ilmaisujen kontekstisidonnaisuuteen eli indeksisyyteen punoutuu refleksiivisyys, jonka Garfinkel ja Sacks (1970) esittävät vuorovaikutustoimintojen olennaisena piirteenä. Tuottamallaan vuorolla puhuja osoittaa tulkintansa siitä, millainen toiminta on meneillään, ja samanaikaisesti vuorolla on seurauksia vuorovaikutuksen etenemiselle – puheenvuoro synnyttää konteksin, jossa se itse esiintyy. Heritage (1996 [1984], 237) kiteyttää tämän sosiaalisten toimintojen ja puheenvuorojen kaksoiskontekstuaalisuudeksi: jokainen toiminto on kontekstin muovaama ja kontekstia uudistava. Näin ollen mitään vuorovaikutuksen yksityiskohtaa ei keskustelunanalyyttisessa tarkastelussa voida sivuuttaa apriorisesti sattumanvaraisena tai merkityksettömänä.

Vuorovaikutuslingvistiikka jakaa monia keskustelunanalyyttisia käsityksiä. Niinpä katsotaan, että osallistujat aktiivisesti synnyttävät vuorovaikutuksen reaaliajassa sen, mitä pidetään kielellisinä yksikköinä, kuten lausekkeen tai lauseen. Yksiköt muotoutuvat vuorovaikutuksen edetessä, ja niiden muotoutumiseen vaikuttavat erilaiset puheen ajalliset ominaisuudet ja rajoitteet, kuten sen ohimenevyys, peruuttamattomuus sekä puhumisen ja puheen vastaanottamisen välinen synkronisaatio (Auer 2009). Kielellinen yksikkö voidaan tuottaa myös yhteistyössä niin, että se jakautuu useammalle puhujalle (esim. Hayashi 2003; Helasvuo 2004). Prosessia, jossa sekä yksittäinen, hetkellinen kielellinen tuotos että pysyvämmäksi kiteytyvä kielioppi syntyvät ja muovautuvat kielen käytön myötä, kutsutaan *emergenssiksi* (ks. Hopper 1987; 2011; Helasvuo 2009; Pekarek Doehler 2011; myös Lindström 2006). Kielellisten käytänteiden emergenssi ja kielenkäytön konteksti, etenkin sen sekventiaalinen konteksti, kuuluvat olennaisesti vuorovaikutuslingvistiseen teoriaan. (Ks. tarkemmin Couper-Kuhlen & Selting 2017.)

Keskeinen vuorovaikutuksen ja kieliopin hetkessä muotoutuva ominaisuus on projektio eli se, millaisia odotuksia ja rajoitteita jonkin elementin tuottaminen asettaa vuoron ja toiminnan etenemiselle (ks. Auer 2005; Auer & Pfänder 2011; Couper-Kuhlen & Selting 2017). Nämä olosuhteet koskevat sekä vuoron tuottajaa että vastaanottajaa. Kielellisten rakenteiden ja toiminnan projektio mahdollistaa muun muassa sen, että keskustelun osallistujat voivat tulkita ja tunnistaa toimintoja jo varhaisessa vaiheessa toisen vuoroa, mikä taas mahdollistaa saumattoman vuoronvaihdon ja yleisesti keskustelun jouhevan etenemisen.[8] Näin ollen kielellisiä rakenteitakaan ei analysoida vain puhuessa syntyneen lopputuloksen eli jonkin tuotetun ilmauksen näkökulmasta vaan ottaen huomioon myös niiden ajallisen ja vuorovaikutuksellisen syntyprosessin (vrt. Kieliopin tutkimus ja kielioppiteoriat, luku 3.6. tässä kirjassa).

## 2.5. Eettisiä kysymyksiä

Keskustelunanalyyttinen tutkimus perustuu tyypillisesti aitojen, ilman koejärjestelyjä tuotettujen keskustelutilanteiden tarkkaan analyysiin. Tarkasteltavien tilanteiden kirjo on laaja aina perheenjäsenten ja ystävien välisistä arkisista keskusteluista institutionaalisiin ja ammatillisiin tilanteisiin. Aineisto on tallennettava, jotta sitä voidaan luotettavasti analysoida.

Tutkimuksessa kannattaa suosia sellaista aineistoa, jossa tutkija itse tai tämän läheiset eivät ole mukana, jotta suhteiden läheisyys ei hämärrä analyysia. Kaikissa tapauksissa jokaiselta tilanteen osanottajalta tarvitaan etukäteen lupa vuorovaikutustilanteen taltiointiin, käsittelyyn ja arkistointiin. Vaikka tutkimuksessa esiinnytään nimettömästi, ääni ja kuva ovat tunnistetietoja, joten aineiston säilyttäminen ja käsittely on järjestettävä niin, että tutkittavien tietosuoja taataan annetun suostumuksen mukaisesti. Tähän sisältyy muun muassa tunnistetietojen muuttaminen litteraatioissa. EU:n uusi tietosuoja-asetus astui voimaan toukokuussa 2018, ja uusia kansallisia käytänteitä ja ohjeistuksia on laadittu. Tutkimuslupiin sisällytettävistä seikoista, lupien muotoilusta ja tutkittavien informointitavasta, aineiston käsittelystä ja turvallisesta säilyttämisestä

saa tietoa yliopistojen tietosuojavastaavilta, tekniseltä henkilökunnalta ja kokeneilta aineistonkerääjiltä.

Arkistoidut vuorovaikutusaineistot mahdollistavat erilaisia tutkimusnäkökulmia, mukaan lukien historiallisen tutkimuksen ja monenlaisen kieltä koskevan tutkimustiedon soveltamisen, jota yhteiskunnassa tarvitaan ja johon kielentutkijoilla on metodisia valmiuksia. Tutkimusaineistojen asianmukainen säilyttäminen ja saatavuus onkin ensiarvoisen tärkeää.

# 3. Tutkimisen käytänteitä ja kohteita

## 3.1. Tutkimusprosessin eteneminen

Tutkimus lähtee liikkeelle joko tarkasteltavan kielellisen rakenteen tai toiminnon taikka vuorovaikutustilanteen valitsemisesta ja aineiston hankkimisesta. Aineiston valinnassa voi painottua tietty toimintaympäristö (työyhteisö, luokkahuone, perheen aamiaishetket), tai sitten voidaan tarkastella jotakin vuorovaikutuksen ilmiötä tai kielellistä rakennetta erilaisten tilanteiden kautta ja pyrkiä yleistyksiin sen tehtävistä. Arkikeskusteluja pidetään lähtökohtaisena aineistona, koska ne ovat monessa mielessä ensisijainen kielenkäytön konteksti (Drew & Heritage 1992; Raevaara, Ruusuvuori & Haakana 2001; Heritage & Clayman 2010).

Tarpeen mukaan voidaan joko käyttää valmiiksi saatavilla olevia aineistoja tai kerätä uusia. Aiemmin koottua puhutun vuorovaikutuksen aineistoa on tarjolla yliopistojen arkistoissa (esim. Helsingin yliopiston Keskusteluntutkimuksen arkisto). Arkikeskustelujen morfosyntaktinen tietokanta Arkisyn on julkisesti käytettävissä Kielipankissa (tietokanta on laadittu Turun yliopiston suomen kielen oppiaineen hankkeessa, ja perusaineisto on Helsingin ja Turun yliopistoista). Kun hankitaan uutta aineistoa, se tallennetaan hyvätasoisin video- ja äänitallentimin. Kameran kuvakulman on tärkeä kattaa sekä yksityiskohdat että ympäröivä tilanne mahdollisimman laajasti, ja usein tarvitaankin useampi kamera (esim. tutkittaessa luokkahuonevuorovaikutusta) – kasvokkaisen

vuorovaikutuksen tutkimukseen taltiointi, jossa kameran kuvakulmassa on vain yksi henkilö, on käytännössä käyttökelvoton. Tutkijan kannattaakin tutustua kuvauspaikan puitteisiin ja suunnitella kuvakulmia ennakkoon. Sopivien tallennusvälineiden ja -formaattien sekä kuvauskulmien valitsemiseksi voi hakea neuvoa asiantuntijoilta. Ennen vuorovaikutuksen tallentamista on hyvä tehdä ainakin jonkin verran etnografista kenttätyötä kuvattavien ja kuvauspaikan toimintojen ymmärtämiseksi, merkiten muistiin tallennustilannetta koskevia seikkoja. (Systemaattisen johdatuksen keskustelunanalyyttisen tutkimuksen aineiston keräämisen eri puoliin esittää Mondada 2012a.)

Aineistonkeruun jälkeen aineisto käsitellään muun muassa nimeämällä tiedostot systemaattisesti ja tekemällä varmuuskopiot sekä kirjaamalla metatietoja (aika ja paikka, kesto, osanottajien kuvaus jne.). Aineistoon tutustumisen voi aloittaa kuuntelemalla tai katsomalla tallenne kokonaisuudessaan ja merkitsemällä muistiin havaintoja. Näin aineiston kokonaisuuteen saa jo alustavan näkökulman. Tämän jälkeen voi ruveta litteroimaan oman kiinnostuksen kohteena olevia kohtia. Puheen litterointiin voidaan käyttää erityisiä ohjelmia, kuten ELAN ja Transana. Litterointitapa ja tarkkuusaste valitaan tutkimuksen tarkoituksen mukaan, ja jokainen litterointitapa on väistämättä valikoiva. Usein alun karkeaa litteraatiota tarkennetaan vasta, kun on päätetty tarkemman analyysin kohteet.

Keskustelunanalyyttisen tutkimuksen tarpeisiin on kehitetty oma litterointitapansa (liite 1), joka kuvaa puheen etenemistä ja tuottamista osanottajien välisen vuorovaikutuksen näkökulmasta; litterointitavan loi alkuaan Gail Jefferson (ks. Seppänen 1997b; Sorjonen & Raevaara 2006; Hepburn & Bolden 2012). Saksalaiset vuorovaikutuslingvistit ja keskustelunanalyytikot ovat laajentaneet jeffersonilaista litterointitapaa kielellisten ja foneettisten seikkojen, erityisesti prosodian esittämiseksi (Selting ja muut 2011). Jotkut tutkijat käyttävät tarkemman foneettisen esittämisen tarpeisiin myös kansainvälistä foneettista aakkostoa (IPA, *International Phonetic Alphabet*)[9] tai suomalais-ugrilaista tarkekirjoitusta (SUT, ks. Sovijärvi & Peltola 1977). Foneettisen analyysin tarpeisiin käytetään Praat-ohjelmaa.[10] Multimodaalista litterointitapaa (liite 2) on kehittänyt erityisesti Lorenza Mondada, jonka litterointisysteemiä käytetään

yhä laajemmin keskustelunanalyyttisessa tutkimuksessa (Mondada 2016; 2018; Harjunpää, Mondada & Svinhufvud 2019). Aineiston esittämistapojen ja itse analyysin yksityiskohtaisuus liittyy pyrkimykseen rajata ja jäsentää tutkittava ilmiö vuorovaikutuksen osallistujien ja heidän havaitun toimintansa ehdoilla huomioiden, että mikä tahansa vuorovaikutuksen elementti voi osoittautua analysoitavan ilmiön kannalta merkitykselliseksi.

Vuorovaikutuksen analyysi sisältää erilaisia vaiheita (ks. esim. Sidnell 2009a; 2009b; 2012; Vatanen 2016). Analyysi voi lähteä liikkeelle jostakin havainnosta, jonka tutkija tekee aineistoa kuunnellessaan ja katsoessaan. Havainnon laajentamiseksi ja havaitun ilmiön kuvaamiseksi muodostetaan esimerkkikokoelmia (engl. *collection*). Aluksi kerätään kaikki tiettyä käytännettä koskevat tapaukset, esimerkiksi kohdat, joissa puhuja ei jatkakaan puheenvuoroaan konjunktiona käytettävän sanan jälkeen (esim. *ja*, *mutta*, *että*) ja vastaanottaja ottaa vuoroon (ks. Koivisto 2011). Kun tapaukset on kerätty, niitä ruvetaan tarkastelemaan yksityiskohtaisemmin ja vertailemaan keskenään: onko kaikilla jotakin yhteistä vai onko kyseessä useampi vuorovaikutusilmiö, jotka voidaan jakaa omiksi kokoelmikseen. Vaihtelemalla työtapaa yksittäisten tapausten analyysista kokoelmien tarkasteluun ja takaisin päästään lopulta yleisempään ilmiön kuvaukseen ja yleistyksiin.

Analyysissa etsitään näyttöä vuorovaikutuksen osallistujien omasta suuntautumisesta havaittuihin käytänteisiin: tarkastellaan muun muassa vuoron suhdetta aiempiin vuoroihin, vuoron muotoilua kokonaisuutena, vaihtoehtoisia muotoiluja eri tapauksissa ja sitä, miten vuoroon vastataan. Havaitusta ilmiöstä näennäisesti poikkeavia tapauksia (engl. *deviant case*) voidaan käyttää havaintojen vahvistamiseen osoittamalla osallistujien itsensä suuntautuminen taustalla olevaan käytänteeseen tai muu tilanteinen poikkeamaa selittävä seikka – muussa tapauksessa joudutaan tarkentamaan itse ilmiön kuvausta (ks. ten Have 2007, 151). Esimerkiksi vierusparin osien eli etu- ja jälkijäsenen (kuten kysymys ja vastaus) kiinteä suhde on osoitettu sillä, että osallistujat selittävät tai paikkaavat jälkijäsenen puuttumista toiminnallaan. Jotkut tutkijat kvantifioivat löydöksiään, toiset taas pitäytyvät laadullisessa tutkimuksessa, joka voi koskea jopa vain yksittäistä tapausta (esim. C. Goodwin 1979; Whalen, Zimmerman & Whalen 1988). Tapausten laadullinen vertailu

kuuluu kuitenkin yleensä keskeisesti tutkimusprosessiin (Haakana, Laakso & Lindström 2009).

Kokoelmien ja vertailun merkitys sekä aineistolähtöisyys näkyvät muissakin tutkimuksen vaiheissa. Keskustelunanalyytikot käyttävät havaintojen muodostamiseen ja testaamiseen yhteisöllistä analyysin muotoa, datasessioita (ks. Peräkylä 1997; Stevanovic & Weiste 2017). Datasessiossa osanottajat tarkastelevat aineistoa yhdessä ja keskustelevat havainnoistaan metodiseen asiantuntemukseensa, aiempaan tutkimukseen ja osin myös intuitioonsa nojaten. Kollektiivinen kriittinen tarkastelu vahvistaa havaintojen ja analyysien luotettavuutta. Saman periaatteen mukaisesti myös tutkimusten julkaisuissa näytetään kattavasti aineistoesimerkkejä, joiden perusteella lukija voi arvioida esitettyjä tuloksia.

## 3.2. Rakenteet, resurssit ja toiminta

Kaikessa tutkimuksessa otetaan huomioon keskustelun jäsennyskeinot: vuorottelu ja vuoron rakenneyksiköt, sekvenssien rakentuminen ja se, miten kohdetoimintoon vastataan ja miten se tulkitaan, sekä puheen etenemisen katkokset ja korjaukset. Ne eivät perustu kasvokkaisissa tilanteissa vain kielellisten vaan myös kehollisten resurssien käyttöön. Siksi keskustelunanalyyttisessa tutkimuksessa käytetäänkin yhä enemmän videoaineistoja.

Tässä jaksossa havainnollistamme kahden videoidun ja litteroidun keskustelukatkelman avulla keskustelunanalyyttista ja vuorovaikutuslingvististä tutkimusotetta. Tarkoituksemme on näyttää esimerkkien avulla, mihin analyysissa saatettaisiin kiinnittää huomiota ja millaisiin kysymyksiin se voisi tarjota vastauksia. Kumpikin esimerkki on institutionaalisesta, vähintään toisen osapuolen työtehtäviin liittyvästä tilanteesta, jossa vuorovaikutuksella on jokin institutionaalinen päämäärä (institutionaalisen vuorovaikutuksen tutkimuksesta ks. Drew & Heritage 1992; Ruusuvuori, Haakana & Raevaara 2001; Heritage & Clayman 2010). Ensimmäinen esimerkkimme on kasvokkainen asiointitilanne kioskista ja toinen kahden yrityksen välillä puhelimitse käyty etämyyntineuvottelu, jossa käytetään jaettua tietokoneruutua.

**Kaksi tapausta**. Ensimmäisessä katkelmassa asiakas tulee kioskille ostamaan päivän lehteä, jonka etusivulla on juttu Suomen euroviisuedustajan heikosta menestyksestä. Katkelma osoittaa kielellisten ja kehollisten resurssien hienovaraisen yhteispelin sinänsä yksinkertaisessa vuorovaikutustapahtumassa. Havainnollistamme katkelman tarkastelussa sitä, miten asiakkaan pyytämä lehti tulee myyjän ja asiakkaan jaetun huomion kohteeksi kielellisesti ja kehollisesti, muun muassa sen kautta, miten demonstratiiviset paikanilmaukset suhteutuvat toiminnan etenemiseen. Käytetyt litterointimerkit ovat artikkelin lopussa liitteenä.

Esimerkki 1. Iltalehti (t228 00:15–00:32)
MYY=myyjä, ASI=asiakas

```
01  ASI: ööh (.) ilta+lehti >ei kai se< teill_ole
    asi              +katsoo ympärilleen-->
02       [l o p]pun’t°.
03  MYY: [joo,h]
04       (0.2)+(0.4)            +(0.1)
    asi    ->+kats. lompakkoaan+kohottaa katsettaan->
05  ASI: eihän+ teill_ole #loppu#nu.
    asi    ->+katsoo lompakkoaan->
06       (0.8)
07  MYY: e:i; ota< (.) >iltalehev vai il+t#a#<
    asi                               ->+katsoo alas laukkuaan->
08  ASI: il[ t a * l e]*hti.
09  MYY:   [>sanomat<.]
    myy          *.....*lukee viivakoodin->
10  MYY: [jaha;
11  ASI: [(joo.)
12       (1.3)*
    myy    ->*
13  MYY: *olikos >sii<nä # ee*+ssä.           +      *
    myy  *...................*käsi tiskin etureunalla*
    asi                       +katsoo tiskille+...->
    kuv                 #Kuva1
```



Kuva 1.

```
14        *(0.2) # +  (0.3) *(0.1) +(0.2)*(0.1) #
     myy  *liu'uttaa kättään*............*tarttuu lehteen->
     asi        ->+katsoo m. kättä+lehteä-->
     kuv        #Kuva2                              #Kuva3
```



Kuva 2.



Kuva 3.

```
15   MYY: >tuossa<.*
     myy        ->*nostaa lehden esiin->
16   ASI: o:nhan siinä +joohh.
     asi            ->+katsoo lompakkoaan, kaivaa rahoja->>
17        (0.4)*(0.7)
     myy     ->*asettaa lehden tiskille->
18   ASI: voi @voi ku* sille #lauralle nyk kävi:#@
     myy              ->*
19   ASI: tuolla [lailla°.
20   MYY:       [mm↓:-m,
```

Katkelmassa toiminta etenee erilaisten vaiheiden kautta pyynnön esittämisestä (r. 1) sen täyttämiseen. Käsittelemme ensiksi osallistujien kielellistä ja kehollista orientaatiota lehden etsimiseen.

Katkelman alussa asiakas osoittaa epävarmaksi, onko lehteä saatavilla: lausekemuotoinen pyyntö *iltalehti* laajentuu kieltomuotoiseksi kysymykseksi, joka hakee vastaukseksen kieltoa (r. 1–2 ja uudestaan r. 5). Vuoroa tuottaessaan asiakas vilkuilee ympärilleen kuin lehteä hakien mutta kohdistaa sitten huomionsa lompakon ja laukun käsittelyyn. Myyjän vastaus vahvistaa lehden saatavuuden (r. 7). Kun myyjä on lukenut Iltalehden kiinteän viivakoodin tiskin pinnasta (r. 9), hän kumartuu tiskin yli asiakkaan puolelle ja kurkottaa kättään kohti lehtipidikettä (kuva 1). Samalla hän tuottaa kysymykseksi muotoillun vuoron *olikos siinä eessä*, jolla osoittaa epävarmaksi, onko mainittuja lehtiä jäljellä pidikkeessä. Paikanilmauksen demonstratiivinen osa *siinä* ilmaisee alueen olevan vastaanottajan havaittavissa ja relevantti tämän toiminnan kannalta (ks. Laury 1997, 77–79). Asiakas osoittaa vastaavaa tulkintaa kohottamalla katseensa tiskin suuntaan. Nähtyään myyjän tiskin yli kurkottavan käden asiakas seuraa katseellaan sen liikettä (r. 13–14, kuva 2), kunnes saavuttaa katseellaan lehden (kuva 3), johon myyjä on juuri tarttumassa. Myyjän puhe ja kehollinen toiminta yhdessä siis kutsuvat asiakkaan kohdistamaan huomionsa edessään olevaan lehtivalikoimaan.

Kun myyjä nappaa lehden, hän ilmaisee sen löytymistä vuorollaan *tuossa* (r. 15). Demonstratiivin valinta osoittaa muuttunutta orientaatiota lehteen ja sen sijaintiin tilassa (vrt. aiempi *siinä*). Myyjä ikään kuin osoittaa juuri havaitsemaansa lehteä asiakkaalle (Laury 1997, 74–75; Etelämäki 2006, 116–117). Asiakas puolestaan vahvistaa *joo*-loppuisella vuorollaan myös havainneensa lehden (r. 16). Tästäkin vuorosta voidaan tarkastella piirteitä, jotka osoittavat asiakkaan uutta tulkintaa ympäristöstään lehden löydyttyä. Liitepartikkeli -*han* ilmaisee äkillistä oivallusta (ISK § 830), joka pohjaa juuri edellä kehollisessa vuorovaikutuksessa syntyneeseen visuaaliseen havaintoon. Paikanilmauksella *siinä* asiakas käsittelee aivan edessään olevaa, nyt yhteisesti tarkasteltua ja tunnettua tilaa myyjän toiminnan alueena. Huomattavaa esimerkissä on siis muun muassa se, miten demonstratiivikantaisten paikanilmausten *siinä*

ja *tuossa* käytöllä jäsennetään vuorovaikutuksellista tilaa ja osallistujien välisen toiminnan etenemistä.

**Kielellisten resurssien tutkimuksesta: demonstratiivipronominit vuorovaikutuksessa**

Suomen demonstratiivipronomineista *tämä*, *tuo* ja *se* on tehty vuorovaikutuksellista tutkimusta, joka täydentää aiempia kielitieteellisiä näkemyksiä osoittamalla demonstratiivien käytön sosiaalisen ja vuorovaikutuksellisen perustan. Demonstratiivit eivät ilmaise vain puhujan ja vastaanottajan etäisyyttä tarkoitteesta vaan jäsentävät osallistujien keskinäisiä vuorovaikutuksellisia asemia ja meneillään olevaa toimintaa, kuten esimerkissä nähdään demonstratiivikantaisen (*se*) paikan adverbin *siinä* käytöstä: vaikka demonstratiivipronominia *se* on usein pidetty etäisyyttä osoittavana, voi puhuja viitata sillä välittömässä läheisyydessäänkin olevaan paikkaan tai tarkoitteeseen, kun käsittelee sitä vastaanottajan toiminnan kohteena.

Laury (1997) on osoittanut tutkimuksessaan, miten pronominia *se* käytetään osoittamaan tarkoitteen tunnettuutta ja että sillä on artikkelimaista käyttöä. Seppäsen (1998) tutkimus viittaamisesta läsnäoleviin henkilöihin taas näyttää, miten viitattaessa demonstratiivipronomineilla läsnäoleviin henkilöihin jäsennetään keskustelun osallistumiskehikkoa ja sen muutoksia. Etelämäki (2006) puolestaan osoittaa, miten *tämä* ja muut demonstratiivit suhteuttavat tarkoitteen tunnistamisen meneillään olevaan toimintaan ja järjestävät taustanaan olevaa keskustelutilannetta. Tutkimuksissa keskustellaan myös siitä, millaisten teoreettisten käsitteiden avulla voidaan parhaiten kuvata demonstratiivien indeksikaalisia suhteita puhekontekstiinsa.

Vuorovaikutuksellisen tilan jäsentäminen toimintojen kielellisessä muotoilussa liittyy myös tilanteen institutionaalisuuteen. Myyjällä ei institutionaalisesta roolistaan huolimatta olekaan tässä tilanteessa ensisijaista pääsyä tuotteeseen, sillä tuote kuuluu asiakkaan itsensä

poimittaviin ja on sijoitettu kioskilla tämän mukaisesti. Asetelma näkyy myyjän vuorossa, jossa hän osoittaa hakevansa asiakkaan pyytämää lehteä paikasta, johon tällä on ensisijainen kehollinen pääsy.

Katkelmasta voitaisiin myös lähteä tutkimaan esimerkiksi affektista asennoitumista sekä rupattelua (engl. *small talk*). Rivin 18 vuoron alussa asiakas ilmaisee affektista suhtautumistaan Lauran heikkoon menestykseen interjektiolla *voi voi* (vrt. Hakulinen & Sorjonen 2012; ks. myös Raevaara & Sorjonen 2006) ja vetoaa sitten huonon uutisen jaettuuteen (*ku sille Lauralle nyt kävi tuolla lailla*)[11]. Myyjä vastaa vuorollaan *mm↓:-m*, jonka prosodista muotoilua kaksihuippuisena, ensimmäisen osan lopussa äännettä venyttäen sävelkulun samalla hetkeksi madaltuessa, voisi tarkastella samanlinjaisen affektin tuottamisen keinona. Asiakkaan voi tässä ajatella esittävän motiivin lehden ostamiselle yhtäältä affektisen merkityksen ja toisaalta tiedon jaettuuden pohjalta. Uutisen mainitseminen tällä tavalla toimii niin ikään resurssina rupattelulle, joka väljentää tilanteen institutionaalisuutta arkisen juttelun suuntaan (Raevaara & Sorjonen 2006; Raevaara 2009; 2011).

Toisessa esimerkissä tuomme esiin, miten keskustelijoiden käyttämiä kielellisiä resursseja voidaan tarkastella suhteessa heidän institutionaalisiin rooleihinsa. Kuten esimerkissä 1 myös esimerkissä 2 on kyseessä osto- ja myyntikonteksti. Konteksti eroaa kuitenkin monin tavoin kioskiympäristöstä: nyt kyseessä on kahden yrityksen välinen myyntineuvottelu. Yritysten välisessä kaupankäynnissä myyjän tehtävä on vakuuttaa asiakas siitä, että tarjottu palvelu olisi tälle vähintään hintansa arvoinen. Asiakas puolestaan usein jarruttelee myyjää ja esittää varauksia, joilla esimerkiksi vähättelee tarvettaan ja pyrkii tinkimään tarjotun palvelun hinnasta (Niemi & Hirvonen 2019). Myyjän ja asiakkaan tuottamia toimintoja voidaan tutkia tarkastelemalla niiden sekventiaalista rakentumista ja kielellistä muotoilua.

Esimerkissä 2 myyntineuvottelu käydään puhelimitse. Osallistujilta siis puuttuu edeltävässä katkelmassa havainnollistettu resurssi toiminnan ymmärrettäväksi tekemisessä: toistensa suoraan havaittavissa oleva kehollinen toiminta. Heillä on kuitenkin käytössään jaettu visuaalinen kanava eli tietokoneruutu, jota myyjä voi käyttää tarjoamansa palvelun esittelyyn tekstin ja kuvien avulla. Myyjä kauppaa esimerkiksi

työajanseurantaan ja työmaapäiväkirjan pitoon käytettävää tietokonesovellusta asiakkaalle, joka edustaa rakennusalalla toimivaa yritystä. On käynyt ilmi, että asiakas osallistuu esimerkiksi taloyhtiöiden remontteihin konsulttina ja osana laajempaa aliurakoitsijoiden joukkoa. Tämän tiedon pohjalta myyjä alkaa kartoittaa asiakkaan tarvetta kauppaamansa palvelun sovellusmahdollisuuksille. Myyjä esittää esimerkin 2 alussa väittämän, joka koskee asiakkaan tietämyksen aluetta (niin kutsuttu B-tilanteen väittämä; Labov 1972a, 254; Labov & Fanshel 1977, 100–101).

Esimerkki 2. (MOV-HAR meet1 13:20–13:50)
M=Myyjä, A=Asiakas

```
01 M:   joo. .hhhhhhhh ↓okei eli teillä ei oo (.) mitään
02      verottajarepo- (0.2)
03      raportointivelvol[lisuuksia (--),                  ]
04 A:                    [ei o ei o meill_ei oo kato semmosia.]
05 M:   j:oo[h.
06 A:       [et (.) et ei oo että et ↑meiän meiän tarve
07      oikeestaan on vaan se että (0.4) et tota noin (0.2)
08      saadaan tämmönen niinku suht reaaliaikanen, (0.4)
09      työmaapäiväpäivä- (.) päiväkirjajärjestelmä ja (0.4)
10      tavallaa niinku niitten (0.6) työmaa- (.) vaiheiden
11      (0.2) dokumentointi ja (0.6) ja,
12 M:   joo-[o.
13 A:       [jakelu.
14      (0.4)
15 A:   tämmöne (.) tämmöne niin se on se meiän tarve,
16      (0.4)
17 M:   ↓okei.
```

Myyjän kieltomuotoinen eksistentiaalilause (r. 1–3; vrt. esimerkin 1 *iltalehti ei kai se teill ole loppun't*) toimii polaarisena kysymyksenä, joka virittää odotuksen asiakkaan vahvistuksesta. Lauseen kieltohakuinen kvanttoripronomini *mitään* ja monikkomuotoinen, partitiivisijainen subjekti (*raportointivelvollisuuksia*) viittaavat tarkemmin rajaamattomaan

luokkaan kaupatun palvelun mahdollisia ominaisuuksia (ISK § 758 ja 1733; Korpela 2007). Asiakkaan vastauksesta voidaan esimerkiksi tarkastella niitä piirteitä, jotka osoittavat asiakkaan suuntautumisen vuorovaikutustilanteen laajempaan kontekstiin eli myyjän pyrkimykseen tehdä kauppaa. Asiakas vahvistaa ensin kieltomuotoisen väittämän ja osoittaa riittävää ymmärrystä myyjän implikoimasta luokasta (*semmosia*, r. 4; Vilkuna 1992 [1980], 132–133; Erringer 1996; ISK § 725). Samalla hän implikoi partikkelilla *kato* (Hakulinen & Seppänen 1992) vetoavansa siihen, että vastaanottaja ymmärtää kieltävän vastauksen. Jatkaessaan vuoroaan ja selvittäessään tarpeitaan asiakas käyttää partikkeleita *oikeestaan* ja *vaan* vähättelemään ja rajaamaan tarvetta (r. 7). Keskustelunanalyyttinen tutkimus voisi tarkastella useampia esimerkkejä analysoimalla sitä, miten asiakas muotoilee vastauksensa niin, että tarjoaa paitsi vastauksen esitettyyn kysymykseen myös suuntautuu tilanteen laajempaan kehykseen eli myyjän rooliin liittyvään pyrkimykseen tehdä kauppaa.

Yksi institutionaalista vuorovaikutusta tarkastelevan keskustelunanalyyttisen tutkimuksen linjoista selvittää, miten keskustelijat omilla toiminnoillaan rakentavat ja tuovat ilmi tilanteeseen ja omiin rooleihinsa liittyviä odotuksia ja velvollisuuksia (esim. Heritage & Clayman 2010, 20–22). Esimerkissä 2 asiakkaan voidaan tulkita tuovan esiin oman ymmärryksensä vuorovaikutustilanteesta juuri sillä, miten hän muotoilee vastauksensa myyjän polaariseen kyllä/ei-kysymykseen: asiakas osoittaa ymmärtäneensä, että myyjän kysymys raportointivelvollisuuksista kartoitti hänen mahdollista tarvettaan tietylle tuotteelle, ja vastauksen laajennus tarjoaa selvityksen näistä tarpeista. Myyjä ottaa partikkelilla *okei* (r. 17) asiakkaan laajan tiedonannon vastaan ymmärrettynä ja riittävänä.

**Toimintojen muotoilu**

Toimintojen muotoilulla tarkoitetaan sitä, miten kieltä, kehoa ja toimintaympäristön resursseja käytetään rakennettaessa esimerkiksi kutsua, valitusta tai samanmielisyyden osoitusta keskustelukumppaneille tunnistettavaksi toiminnoksi. Viime vuosina paljon huomiota ovat saaneet esimerkiksi toimintajakson aloittavan pyynnön (Curl & Drew 2008; Drew

& Couper-Kuhlen 2014; Rossi 2015), direktiivin (Sorjonen, Raevaara & Couper-Kuhlen 2017; ks. myös Lauranto 2015) ja kehun esittäminen (Golato 2005; Etelämäki, Haakana & Halonen 2013), mutta tutkimuksissa on käsitelty myös responsiivisten toimintojen muotoilua (Sorjonen 2001a; Hakulinen & Sorjonen 2009; Niemi 2015; Thompson, Fox & Couper-Kuhlen 2015; Koivisto 2017) ja osoitettu hienovaraisia eroja toisistaan muodoltaan vain minimaalisesti eroavien toimintojen välillä.

Toimintojen muotoiluun liittyy yleisemmän tason ilmiöitä, kuten preferenssijäsennys: tietyille toiminnoille voidaan osoittaa odotuksenmukainen eli preferoitu ja epäodotuksenmukainen eli preferoimaton responssi – preferoimattoman toiminnan muotoiluun liittyy tyypillisesti vuoron viivästyminen tai esimerkiksi selonteon tuottaminen (preferenssistä keskustelun eri ilmiöiden tasolla ks. Pomerantz & Heritage 2013; myös Raevaara 2011). Osallistujat ilmaisevat toimintojen muotoilulla myös episteemistä eli tiedollista asemaa ja asennoitumista (*epistemic status/stance*, Heritage 2012a; 2012b). Episteemisyys onkin noussut 2000-luvulla merkittäväksi teemaksi keskustelunanalyysissa.

Toimintojen tutkimuksessa, kuten keskustelunanalyysissa ylipäänsä, kehollisen toiminnan tarkastelu ja ympäristön fyysisten resurssien huomiointi on vahvistunut kielellisen ulottuvuuden rinnalla (Haddington & Kääntä 2011; Nevile, Haddington, Heinemann & Rauniomaa 2014).

**Vertaileva ja kieltenvälinen tutkimus**. Keskustelunanalyyttinen tutkimus on perusolemukseltaan vertailevaa. Tällä tarkoitetaan sitä, että kun kiinnostuksen kohteena olevasta vuorovaikutuskäytänteestä, toiminnosta tai toimintakokonaisuudesta on kerätty alustava kokoelma samankaltaisia esiintymiä, niitä aletaan tarkastella suhteessa toisiinsa. Koottujen tapausten piirteitä vertailemalla määritetään, onko kyse samasta ilmiöstä vai pitäisikö tapauksia jakaa uusiin kokoelmiin, mikä esiintymiä yhdistää ja mistä yleisemmästä ilmiöstä on kyse (esim. Haakana, Laakso & Lindström 2009; ks. Schegloff 1968 vertailun tärkeydestä ja poikkeustapauksien käsittelystä).

Vuorovaikutuskäytänteitä ja toimintoja vertaillaan erilaisista näkökulmista myös yksittäisiä kokoelmia laajemmin (johdatuksen vertailun

muotoihin esittävät Haakana, Laakso & Lindström 2009; ks. myös esim. Drew 2003; Sidnell 2009a; 2009b). Arkisen perheen ja ystävien välisen keskustelun ja jonkin institutionaalisen tilannetyypin toimintojen ero tulee usein selvästi esiin siinä, miten osanottajat käsittelevät toistensa toimintoja, kuten seuraavissa kahdessa vastauksessa samalla tavalla muotoiltuun kysymykseen:

Esimerkki 3. (Korpela 2007, 18–20)

```
a.
A:  mitäs kuuluu?
B:  no kiitos siinähän se hurisee minä tässä huokailen vaan
    ku .hh on vähän semmonen niin eh .hh kummallinen olo
    vaan ei meinaa päästä käyntiin nyth oik(h)een, he he .hhh

b.
A:  mitäs kuuluu?
B:  no nyt on >semmone vaiva ollu vissii kesästä lähtie<,
    et mua painaa, (0.2) tähä.
```

Vuoronsa muotoilulla kysymyksen vastaanottajat ilmaisevat suuntautumisensa siihen, millaisessa tilanteensa he ovat ja mitä kysyjä heiltä odottaa (tuttava vastaukseksi kuulumisten kysymiseen; potilas vastaukseksi vastaanoton aloittavaan kysymykseen). Samoin erilaisten institutionaalisten vuorovaikutustilanteiden vertaaminen tuo tarkemmin näkyviin vuorovaikutuskäytänteiden samankaltaisuudet ja erot tilanteesta toiseen sekä sen, millaisiin tilanteen ominaisuuksiin käytänteet liittyvät. Esimerkiksi erot tavoissa, joilla terveyskeskuslääkäri ja homeopaatti kysyvät potilaan elämäntavoista (syömis- ja juomistottumuksista) ja potilaat vastaavat kysymyksiin, kertovat ammattilaisten sairautta ja hyvinvointia koskevista teorioista ja tavoitteista sekä kummankin osapuolen suuntautumisesta muihin vastaanottotilanteen piirteisiin (Lindfors & Raevaara 2005).

Kiinnostava kysymys on vuorovaikutuskäytänteiden samankaltaisuus ja vaihtelu eri kielten ja kulttuurien välillä: mikä on samaa

kieliyhteisöstä ja kulttuurista toiseen ja mikä vaihtelee. Keskustelunanalyyttista tutkimusta yhdistää kiinnostus yleisiin vuorovaikutuksen piirteisiin ja tarpeisiin, joihin puhujat luovat ratkaisuja kunkin kielen omien resurssien ja sosiokulttuuristen yhteisöjen omien rakenteiden puitteissa. Keskustelu metodisista valinnoista ja perusoletuksista tällä alueella on vilkasta. Lähestymistavat vaihtelevat keskustelunanalyyttisen tutkimuksen pohjalta laadittujen koodausten ja kvantifioinnin käyttämisestä (esim. Enfield, Stivers & Levinson 2010, kysymys–responssi-jaksot 10 kielessä) laadullisiin tutkimuksiin, joissa eri kielten tutkijat tarkastelevat samaa aihepiiriä oman kielensä näkökulmasta (esim. Sorjonen, Raevaara & Couper-Kuhlen 2017, imperatiivin sisältävät vuorot 10 kielessä).

Kieltenvälisten tutkimusten painopiste vaihtelee selvemmin kielellisestä fokuksesta (esim. kielioppi, kieltomuotoiset kognitiiviset verbit ja episteemisyys, Lindström, Maschler & Pekarek Doehler 2016) enemmän toiminnalliseen fokukseen (esim. kognitiivisen muutoksen ilmaiseminen, Heinemann & Koivisto 2016; kutsun esittäminen, Margutti, Tainio, Drew & Traverso 2018). Kahden kielen käytänteitä on vertailtu esimerkiksi keskustelijoiden välisen työnjaon ilmaisemisessa (sinä teet x niin minä teen y -rakennetyyppi), pyyntöjen esittämisessä ja päällekkäispuhunnassa (Vatanen 2014; Couper-Kuhlen & Etelämäki 2014; Zinken 2016). Saman kielen eri maissa puhuttavissa kielimuodoissa puolestaan on vertailtu esimerkiksi puhuttelukäytänteitä (ruotsin- ja suomenruotsi; Norrby, Wide, Nilsson & Lindström 2018).

**Tilanteiden ja osallistujien moninaisuus, soveltava tutkimus**. Vuorovaikutuksen jäsennyskeinoja valaisevat myös tilanteet, joissa osallistujilla on toisistaan merkittävästi tiettyjen piirteiden suhteen eroavat kielelliset tai vuorovaikutukselliset resurssit. Epäsymmetristen tilanteiden tarkastelu tuo korostetusti näkyviin niin yhteisymmärryksen rajoitteet kuin sen rakentamisen keinot. Klassikoksi ovat muodostuneet Charles Goodwinin (esim. 1995; 2007) tutkimukset siitä, miten afaattinen osallistuja rakentaa kolmea sanaa (*and, yes, no*) hyödyntämällä toimintaansa muiden tuottaman puheen varaan.

Vuorovaikutusresurssien epätasainen jakautuminen ja toisten puheen varaan rakentaminen tulevat esiin myös lasten ja aikuisten

kielenoppimisessa, toisen kielen käytössä (Gardner & Wagner 2004; Kurhila 2006; Hall, Hellermann & Pekarek Doehler 2011; Savijärvi 2011; Merke 2016; Kotilainen, Kurhila & Kalliokoski 2019; Lilja & Tapaninen 2019) ja monikielisessä vuorovaikutuksessa (Bolden 2012; Frick 2013; Henricson 2013; Harjunpää 2017; Paananen 2017; Plejert, Lindholm & Schrauf 2017). Näiden alojen keskustelunanalyyttisella tutkimuksella on yhteys (vuorovaikutus)sosiolingvistiikkaan, ja alat risteävät erityisesti koodinvaihdon ja muiden monikielisten käytänteiden tarkastelussa (esim. Auer 1998; Nilep 2006; Filipi & Markee 2018; ks. myös Sosiolingvistiikka; Kielikontaktien tutkimus, luku 3.2. tk.). Painotukset ovat kuitenkin erilaisia: kielen variaation, muutoksen, kielenvaihdon ja kieli-ideologioiden kaltaisten teemojen sijaan keskustelunanalyyttinen tutkimus kohdistuu tyypillisesti siihen, miten toiminta jäsentyy paikallisesti moninaisten kielellisten resurssien avulla. Usean kielen käyttöä samassa tilanteessa voi lähestyä esimerkiksi sen kannalta, kuinka vuorovaikutuksen kielivalinta tehdään osana tilanteeseen kuuluvaa toimintoa (tervehtiminen asioinnin alussa, Hazel 2015; Mondada 2018) tai miten kielten käyttö jäsentää osallistumista monenkeskisessä tilanteessa (kokousten kulku monikielisillä työpaikoilla, Mondada 2012b; Traverso 2012). Olennaista on myös sen osoittaminen, mitkä seikat eivät palaudu kieliasetelmaan sinänsä vaan muuhun toiminnan jäsentymiseen (Kurhila 2006, 221–225).

### Kielierojen paikkaamisen keinot ja toimintakonteksti

Epäsymmetrisissä monikielisissä tilanteissa paikataan yhteisen kielen puuttumista esimerkiksi korjausjäsennyksen ja kääntämisen keinoin. Savijärvi (2011) on tutkinut, miten suomenkieliset lapset oppivat ruotsia toisena kielenä kielikylpypäiväkodissa. Tutkimuksen pitkittäinen näkökulma tuo näkyviin sen, miten keinot toimia epäsymmetrisessä tilanteessa muotoutuvat yhteisen toiminnan edetessä ja kielitaidon karttuessa. Ensimmäisten kuukausien aikana lapset tukeutuvat kielen sijasta muuhun tilanteen ja toiminnan ymmärrettävyyteen. He alkavat tehdä korjausaloitteita eli osoittaa ymmärrysongelmia vasta siinä vaiheessa,

kun jo ymmärtävät ruotsinkielisten ohjaajien puhetta. Tarkastelemalla monikielisiä tilanteita multimodaalisesta näkökulmasta voidaankin tuoda näkyviin piirteitä, joihin osallistujien vuorovaikutuksellinen toiminta nojaa tietyn kielen osaamisen lisäksi. Harjunpään (2017) tutkimus suullisesta kääntämisestä arkikeskusteluissa muun muassa osoittaa, että käännösvuorojen tuottamista voi ohjata kehollinen neuvottelu sen osallistujan osallistujaroolista, jonka ymmärrystä kääntämällä edesautetaan. Käännösvuorot poikkeavat usein muodoltaan ja sisällöltään käännettävästä vuorosta, sillä käännösvuorot sovitetaan erilaisin käytäntein meneillään olevaan toimintaan, vuoron sekventiaaliseen asemaan ja osallistumiskehikon muutoksiin. Tutkimukset siis näyttävät, miten osallistujien tapa käsitellä tietyn kielistä puhetta ja sen ymmärtämistä kytkeytyy muuhun vuorovaikutuksen ja toiminnan jäsentymiseen.

Yleisemmin tutkimuksen avulla voidaan täsmentää ja muokata ymmärrystä kielellisen ja vuorovaikutuksellisen epäsymmetrian luonteesta ja seurauksista. Niin sanottujen epätyypillisten vuorovaikutustilanteiden tutkimus on otollinen kenttä soveltavalle tutkimukselle, sillä tuloksia on mahdollista hyödyntää eri instituutioiden toiminnan kehittämisessä. Käytännön tarpeista on noussut kiinnostus erityisesti vuorovaikutukselle haasteita tuottaviin afasiaan, dementiaan ja kuulo-ongelmiin, joiden hoito- ja kuntoutuskäytänteiden kehittämiseksi on tutkittu etenkin viime aikoina vuorovaikutusta arkisissa tilanteissa (esim. Leskelä & Lindholm 2012; Pajo 2013; Laakso 2005; 2015; Stevanovic & Lindholm 2016, luku II).

Soveltavan tutkimuksen ominaispiirre on, että tutkijat tekevät yhteistyötä tutkittujen yhteisöjen ja niiden toimijoiden kanssa. Esimerkiksi Kelan puhelinpalvelun vuorovaikutuksen tutkimuksessa aihepiirit, joita tutkimukset koskivat (esim. puhelun aloitus, asiakkaan kuunteleminen ja siirtyminen asian käsittelyyn), valittiin tutkijoiden ja Kelan toimijoiden yhteistyönä. Myös prosessin aikana pidettiin yhteisiä palaute- ja ideointisessioita ja lopuksi julkaistiin laaja raportti (Raevaara, Sorjonen & Lappalainen 2013), joka sisälsi kiteytyksen tutkimuksessa selville saaduista hyvistä käytänteistä. Soveltava tutkimus on kohdistunut myös muun muassa sosiaali- ja terveydenhuollon tilanteisiin, kuten yleislääkärin vastaanottoihin ja psykoterapiatilanteisiin (Sorjonen, Peräkylä

& Eskola 2001; Tiitinen 2015; Weiste & Voutilainen 2016; Lindholm & Leskelä 2018; Paananen 2019) ja työpaikan sisäisiin keskusteluihin (esim. Asmuß & Svennevig 2009; Lehtinen & Pälli 2016; Nissi & Lehtinen 2016; Mikkola & Lehtinen 2019).

Tutkimusta on sovellettu toiminnan kehittämiseen myös erilaisissa oppimisympäristöissä, erityisesti kielenoppimisessa (esim. Kurhila & Kotilainen 2017; ks. myös Kieliopin tutkimus ja kielioppiteoriat tk.). Uusi tutkimuksen sovellusalue on yritysmarkkinointi, jossa on tutkittu muun muassa vapautuneen tunnelman ja asiakasluottamuksen rakentamista kahden yrityksen välisessä myyntineuvottelussa (Kaski, Niemi & Pullins 2018; Humă, Stokoe & Sikveland 2019; Niemi & Hirvonen 2019). Soveltava näkökulma tuo perustutkimuksen tutkijoiden, yhteiskunnan toimijoiden ja kansalaisten käyttöön.

# 4. Yhteenveto ja tulevaisuuden näkymiä

Keskustelunanalyyttisen tutkimuksen ytimessä on tavoite ymmärtää sosiaalisen vuorovaikutuksen toiminnallista jäsentymistä ja keinoja, joilla keskustelun osanottajat tuottavat ja tulkitsevat sosiaalista toimintaa. Puhetta tarkastellaan aina suhteessa muihin vuorovaikutuksen resursseihin, vuorotteluun ja vuorojen sekventiaaliseen kontekstiin sekä näiden paikalliseen muotoutumiseen. Vuorovaikutuslingvistiikassa keskitytään samoista lähtökohdista tutkimaan kielen rakenteita ja niiden käytön vuorovaikutuksellista perustaa. Keskusteluntutkijan ja vuorovaikutuslingvistin analyyttiseen työkalupakkiin sisältyy käsitys keskustelun yleisistä perusjäsennyksistä, joiden avulla saadaan esiin ja voidaan todentaa keskustelun osanottajien tilanteessa jakamia vuorovaikutuskäytänteitä. Toistuvia käytänteitä havainnoimalla puolestaan kasvatetaan yleistettävää ja eri tavoin sovellettavaa tietoa siitä, kuinka keskusteluvuorovaikutus jäsentyy reaaliaikaisesti.

Keskustelunanalyytikot ja vuorovaikutuslingvistit tekevät yhteistyötä eri alojen tutkijoiden kanssa, ja metodia hyödynnetään useilla aloilla.

Aktiivista yhteistyötä on myös muiden kielentutkimuksen suuntausten kanssa, kuten Suomessa kognitiivisen kielitieteen ja konstruktiokieliopin edustajien kanssa (esim. Etelämäki, Herlin, Jaakola & Visapää 2009; Seppänen & Herlin 2009; Helasvuo & Huumo 2010; Haddington, Jantunen & Sivonen 2011; Etelämäki, Herlin, Möttönen & Visapää 2016). Tutkimusyhteistyötä on tehty myös esimerkisi sosiolingvistien (esim. Sorjonen, Rouhikoski & Lehtonen 2015) ja kirjallisuudentutkijoiden (esim. Koivisto & Nykänen 2013; 2016) ja laajasti sosiaalitieteiden edustajien kanssa.

Keskustelunanalyysi laajentuu yhtäältä kattamaan kielellistä ja kulttuurista vaihtelua ja erilaisten vuorovaikutustilanteiden kirjoa. Toisaalta se haastaa entistä voimakkaammin kieleen rajoittuvan tarkastelun riittävyyden vuorovaikutuksen ja jopa kieliopin tutkimuksessa (esim. Keevallik 2018). Tämä valaisee entistä selvemmin kielen ja kehollisuuden suhdetta ja sitä, millä tavalla ja missä määrin kieli on läsnä erilaisissa vuorovaikutustilanteissa. Muista näkökulmista yleistymässä on muun muassa pitkittäinen keskustelunanalyyttinen tutkimus, jossa tarkastellaan vuorovaikutuskäytänteiden ja kielenkäytön muuttumista tietyllä aikavälillä (Deppermann 2018; Pekarek Doehler, Wagner & González-Martínez 2018). Lisäksi teknologiavälitteinen vuorovaikutus (pikaviestimet ym.) ja muu teknologian käyttö arjessa ja työelämässä on nousemassa merkittäväksi keskustelunanalyysin sovellusalueeksi (Giles, Stommel, Paulus, Lester & Reed 2015; Stommel, Paulus & Atkins 2017; Meredith 2019).

Keskustelunanalyysin metodia sovellettaneen jatkossa yhä useammin erilaisissa työelämän vuorovaikutuksen konteksteissa, muita tieteenaloja edustavien tutkimusten ja tutkijoiden rinnalla (esim. Kaski, Niemi & Pullins 2018). Tutkijat ja tutkittavat saavat myös uudenlaisia rooleja erilaisten toimintatutkimusten myötä. Monialaisissa työryhmissä voi työskennellä eri alojen tutkijoiden lisäksi esimerkiksi toimittajia ja taiteilijoita. Aineistopohja voi laajeta niin, että tyypillisen vuorovaikutusaineiston lisäksi hankkeissa käsitellään vaikkapa haastatteluja ja tekstejä. Monialaisissa työryhmissä syntyy myös uudenlaisia tapoja kertoa hankkeiden työstä ja tuloksista.

**Aiheesta lisää:**

Clift, Rebecca. 2016. *Conversation Analysis*. Cambridge: Cambridge University Press.

Couper-Kuhlen, Elizabeth & Selting, Margret. 2017. *Interactional Linguistics. Studying Language in Social Interaction*. Cambridge: Cambridge University Press.

Sidnell, Jack & Stivers, Tanya (toim.). 2012. *The Handbook of Conversation Analysis*. Chichester: Wiley-Blackwell.

Stevanovic, Melisa & Lindholm, Camilla (toim.). 2016. *Keskustelunanalyysi. Kuinka tutkia sosiaalista toimintaa ja vuorovaikutusta*. Tampere: Vastapaino.

Tainio, Liisa. 1997. *Keskustelunanalyysin perusteet*. Tampere: Vastapaino.

## VIITTEET

1 Keskustelunanalyysin syntyaika oli angloamerikkalaisessa tutkimuksessa aikaa, jolloin liikehdintää tapahtui useilla tieteenaloilla ja yhteiskunnasta, kielestä ja sosiaalisesta vuorovaikutuksesta eri tavoin kiinnostuneet tutkijat – sosiolingvistit, lingvistisen antropologian edustajat sekä etnometodologit ja keskustelunanalyytikot – olivat tiiviissä yhteistyössä (Gumperz & Hymes 1972; Bauman & Sherzer 1974).

2 Huomionarvoista on, että artikkeleista kaksi, Sacks, Schegloff & Jefferson (1974) ja Schegloff, Jefferson & Sacks (1977), julkaistiin pohjoisamerikkalaisen kielentutkimuksen lippulaivajulkaisussa *Language*.

3 Levinson luki diskurssianalyysiin tekstikieliopin tutkijat ja diskurssia puheaktiteorian pohjalta tutkivat (esim. Sinclair & Coulthard 1975; Labov & Fanshel 1977).

4 Ryhmään kuuluivat alussa Pia Purra, Liisa Raevaara, Eeva-Leena Seppänen, Marja-Leena Sorjonen ja Liisa Tainio sekä hieman myöhemmin myös Sara Routarinne.

5 Jefferson oli toimittamassa Sacksin luentoja kirjaksi ja lähetti luennot monisteina Hakuliselle sitä vastaan, että sai ne takaisin ylimääräisten kopioiden kanssa.

6 Kuten Levison myös Hakulinen (1989) luki diskurssianalyysiin tekstikieliopin ja puheaktiteorian pohjalta kehitetyn diskurssintutkimuksen. Myöhemmin, vuonna 1997 ilmestyneessä oppikirjassa suomalaiset keskustelunanalyytikot erottivat termit *keskustelunanalyysi* ja *keskusteluntutkimus*. Ensin mainitulla viitattiin etnometodologiseen keskustelunanalyysiin (engl. *conversation analysis*, CA), jälkimmäinen taas oli kattotermi,

joka viittasi millaiseen tahansa tapaan tutkia keskustelupuhetta. (Tainio 1997, 10; ks. myös Tieteen termipankki s.v. keskustelunanalyysi.)

7 Span konferenssin järjestivät Saksassa työskennelleet Margret Selting ja Elizabeth Couper-Kuhlen, Helsingin konferenssin Auli Hakulinen ja Englannissa työskennellyt Bill Wells.

8 Kielten välillä on kuitenkin eroja siinä, missä vaiheessa erilaisia kielen rakenteita ennakoidaan (projisoidaan) (ks. Couper-Kuhlen & Selting 2017).

9 https://www.internationalphoneticassociation.org/content/full-ipa-chart.

10 Praat on digitaalinen puheen analyysiin tarkoitettu ilmaisohjelma. Sen kehittivät 1992 ja sitä päivittävät Paul Boersma ja David Weenik Amsterdamin yliopistossa (Institute of Phonetic Sciences). Ohjelman voi ladata osoitteesta http://www.fon.hum.uva.nl/praat/.

11 Lehden etusivun uutisena on juuri päättynyt Eurovision laulukilpailu ja Suomen edustajan Laura Voutilaisen sijoitus kilpailussa.

## KIRJALLISUUS


Asmuß, Birte & Svennevig, Jan. 2009. Meeting talk: An introduction. *Journal of Business Communication* 46:1, 3–22.

Auer, Peter (toim.). 1998. *Code-switching in Conversation: Language, Interaction and Identity*. London: Routledge.

Auer, Peter. 2005. Projection in interaction and projection in grammar. *Text* 25:1, 7–36.

Auer, Peter. 2009. On-line syntax: Thoughts on the temporality of spoken language. *Language Sciences* 31:1, 1–13.

Auer, Peter & Pfänder, Stefan (toim.). 2011. *Constructions: Emerging and Emergent*. Berlin: de Gruyter.

Bauman, Richard & Sherzer, Joel (toim.). 1974. *Explorations in the Ethnography of Speaking*. Cambridge: Cambridge University Press.

Bolden, Galina. 2012. Across languages and cultures: Brokering problems of understanding in conversational repair. *Language in Society* 41, 97–121.

Clift, Rebecca. 2016. *Conversation Analysis*. Cambridge: Cambridge University Press.

Couper-Kuhlen, Elizabeth. 2018. Finding a place for body movement in grammar. *Research on Language and Social Interaction* 51:1, 22–25.

Couper-Kuhlen, Elizabeth & Etelämaki, Marja. 2014. On divisions of labor in request and offer environments. Julkaisussa: Drew, Paul & Couper-Kuhlen, Elizabeth (toim.) *Requesting in Social Interaction*. Amsterdam & Philadelphia: John Benjamins, 115–144.

Couper-Kuhlen, Elizabeth & Ford, Cecilia (toim.). 2004. *Sound Patterns in Interaction. Cross-linguistic Studies from Conversation*. Amsterdam: John Benjamins.

Couper-Kuhlen, Elizabeth & Selting, Margret (toim.). 1996. *Prosody in Conversation: Interactional Studies*. Cambridge: Cambridge University Press.

Couper-Kuhlen, Elizabeth & Selting, Margret. 2001. Introducing interactional linguistics. Julkaisussa: Selting, Margret & Couper-Kuhlen, Elizabeth (toim.) *Studies in Interactional Linguistics*. Amsterdam: John Benjamins, 1–22.

Couper-Kuhlen, Elizabeth & Selting, Margret. 2017. *Interactional Linguistics. Studying Language in Social Interaction*. Cambridge: Cambridge University Press.

Couper-Kuhlen, Elizabeth & Thompson, Sandra. 2000. Concessive patterns in conversation. Julkaisussa: Couper-Kuhlen, Elizabeth & Kortmann, Bernd (toim.) *Cause, Concession, Condition, and Contrast: Cognitive and Discourse Perspectives*. Berlin: Mouton de Gruyter, 381–410.

Curl, Traci & Drew, Paul. 2008. Contingency and action: A comparison of two forms of requesting. *Research on Language and Social Interaction* 41:2, 129–153.

Deppermann, Arnulf. 2018. Changes in turn-design over interactional histories: The case of instructions in driving lessons. Julkaisussa: Deppermann, Arnulf & Streeck, Jürgen (toim.) *Time in Embodied Interaction*. Amsterdam: John Benjamins Publishing Company, 293–324.

Drew, Paul. 2003. Comparative analysis of talk-in-interaction in different institutional settings: A sketch. Julkaisussa: Glenn, Philip, LeBaron, Curtis & Mandelbaum, Jenny (toim.) *Studies in Language and Social Interaction*. New York (NY): Routledge, 293–308.

Drew, Paul & Heritage, John (toim.). 1992. *Talk at Work. Interaction in Institutional Settings*. Cambridge: Cambridge University Press.

Drew, Paul & Couper-Kuhlen, Elizabeth (toim.). 2014. *Requesting in Social Interaction*. Amsterdam: John Benjamins.

Enfield, Nick E., Stivers, Tanya & Levinson, Stephen C. (toim.). 2010. Question–response sequences in conversation across ten languages. Erikoisnumero. *Journal of Pragmatics* 42:10, 2615–2860.

Erringer, Anu. 1996. The Function of Demonstrative Adjectives *semmoinen*, *tämmöinen* and *tuommoinen* in Finnish Conversations. Pro gradu -tutkielma. Department of Linguistics. University of Colorado.

Etelämäki, Marja. 2006. *Toiminta ja tarkoite. Tutkimus suomen pronominista* tämä. Helsinki: Suomalaisen Kirjallisuuden Seura.

Etelämäki, Marja, Haakana, Markku & Halonen, Mia. 2013. Keskustelukumppanin kehuminen suomalaisessa keskustelussa. *Virittäjä* 117, 460–493.

Etelämäki, Marja, Herlin, Ilona, Jaakola, Minna & Visapää, Laura. 2009. Kielioppi käsitteistyksenä ja toimintana: Kognitiivista kielioppia ja keskustelunanalyysia yhdistämässä. *Virittäjä* 113, 162–187.

Etelämäki, Marja, Herlin, Ilona, Möttönen, Tapani & Visapää, Laura (toim.). 2016. Discourse, grammar and intersubjectivity. Erikoisnumero. *Nordic Journal of Linguistics* 39:2, 101–229.

Filipi, Anna & Markee, Numa (toim.). 2018. *Conversation Analysis and Language Alternation. Capturing Transitions in the Classroom*. Amsterdam: John Benjamins Publishing Company.

French, Peter & Local, John. 1983. Turn competitive incomings. *Journal of Pragmatics* 7, 701–715.

Frick, Maria. 2013. *Emergent Bilingual Constructions. Finnish-Estonian Codeswitching in Interaction*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-9019-6.

Gardner, Rod & Wagner, Johannes (toim.). 2004. *Second Language Conversations*. London: Continuum.

Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Malden (MA): Blackwell Publishers Inc.
Garfinkel, Harold & Sacks, Harvey. 1970. On formal structures of practical actions. Julkaisussa: McKinney, John & Tiryakian, Edward (toim.) *Theoretical Sociology*. New York (NY): Appleton Century Crofts, 338–366.
Giles, David, Stommel, Wyke, Paulus, Trena, Lester, Jessica & Reed, Darren. 2015. Microanalysis of online data: The methodological development of ”digital CA”. *Discourse, Context & Media* 7, 4–51.
Goffman, Erving. 1967. *Interaction Ritual: Essays on Face-to-face Behavior*. New York (NY): Anchor Books.
Goffman, Erving. 1983. The interaction order. *American Sociological Review* 48:1, 1–17.
Goffman, Erving. 2012. *Vuorovaikutuksen sosiologia*. Suomentanut Kaisa Koskinen. Tampere: Vastapaino.
Golato, Andrea. 2005. *Compliments and Compliment Responses: Grammatical Structure and Sequential Organization*. Amsterdam: John Benjamins.
Goodwin, Charles. 1979. The interactive construction of a sentence in natural conversation. Julkaisussa: Psathas, George (toim.) *Everyday Language: Studies in Ethnomethodology*. New York (NY): Irvington Press, 97–121.
Goodwin, Charles. 1981. *Conversational Organization: Interaction Between Speakers and Hearers*. New York (NY): Academic Press.
Goodwin, Charles. 1995. Co-constructing meaning in conversations with an aphasic man. *Research on Language and Social Interaction* 28:3, 233–260.
Goodwin, Charles. 2007. Interactive footing. Julkaisussa: Holt, Elizabeth & Clift, Rebecca (toim.) *Reporting Talk: Reported Speech in Interaction*. Cambridge: Cambridge University Press, 16–46.
Goodwin, Charles. 2017. *Co-operative Action*. Cambridge: Cambridge University Press.
Goodwin, Charles & Goodwin, Marjorie Harness. 2004. Participation. Julkaisussa: Duranti, Alessandro (toim.) *A Companion to Linguistic Anthropology*. Oxford: Blackwell, 222–244.
Goodwin, Marjorie Harness. 1980. Processes of mutual monitoring implicated in the organization of description sequences. *Sociological Inquiry* 50, 303–317.
Goodwin, Marjorie Harness & Cekaite, Asta. 2018. *Embodied Family Coreography. Practices of Control, Care and Mundane Creativity*. Oxford: Routledge.
Gumperz, John & Hymes, Dell (toim.). 1972. *Directions in Sociolinguistics: The Ethnography of Speaking*. New York (NY): Holt, Rinehart and Winston.
Haakana, Markku. 2008. Kieli toimintana. Keskustelunanalyysin näkökulma. Julkaisussa: Onikki-Rantajääskö, Tiina & Siiroinen, Mari (toim.) *Kieltä kohti*. Helsinki: Otava, 86–103.
Haakana, Markku, Laakso, Minna & Lindström, Jan. 2009. Introduction: Comparative dimensions of talk in interaction. Julkaisussa: Haakana, Markku, Laakso, Minna & Lindström, Jan (toim.) *Talk in Interaction: Comparative Dimension*s. Helsinki: Finnish Literature Society, 15–47.
Haakana, Markku, Kurhila, Salla, Lilja, Niina & Savijärvi, Marjo. 2016. Kuka, mitä, häh? Korjausaloitteet suomalaisessa arkikeskustelussa. *Virittäjä* 120:2, 255–293.
Haddington, Pentti, Jantunen, Jarmo & Sivonen, Jari. 2011. Language and affect: go-say and come-say constructions in Finnish. *SKY Journal of Linguistics* 24, 75–117.

Haddington, Pentti, Keisanen, Tiina, Mondada, Lorenza & Nevile, Maurice (toim.). 2014. *Multiactivity in Social Interaction: Beyond Multitasking*. Amsterdam: John Benjamins.

Haddington, Pentti & Kääntä, Leila. 2011. *Kieli, keho ja vuorovaikutus. Multimodaalinen näkökulma sosiaaliseen toimintaan*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Haddington, Pentti & Rauniomaa, Mirka. 2011. Technologies, multitasking and driving: Attending to and preparing for a mobile phone conversation in the car. *Human Communication Research* 37, 223–254.

Hakulinen, Auli (toim.). 1989. *Kieli 4: Suomalaisen keskustelun keinoja 1*. Helsinki: Helsingin yliopisto.

Hakulinen, Auli. 2001a. Eräistä diskurssipartikkelin *kyl(lä)* käyttötavoista keskustelussa. Julkaisussa: Hakulinen, Auli *Lukemisto. Kirjoituksia kolmelta vuosikymmeneltä*. Toimittaneet Lea Laitinen, Pirkko Nuolijärvi, Marja-Leena Sorjonen & Maria Vilkuna. Englanninkielinen alkuperäisartikkeli 2001. Suomentanut Marja-Leena Sorjonen. Helsinki: Suomalaisen Kirjallisuuden Seura, 412–434.

Hakulinen, Auli. 2001b. Minimal and non-minimal answers to yes-no questions. *Pragmatics* 11:1, 1–15.

Hakulinen, Auli & Seppänen, Eeva-Leena. 1992. Finnish *kato*: From verb to particle. *Journal of Pragmatics* 18:6, 527–549.

Hakulinen, Auli & Sorjonen, Marja-Leena. 1993. Research on spoken interaction in Finland. *Research on Language and Social Interaction* 26:4, 409–426.

Hakulinen, Auli & Sorjonen, Marja-Leena. 2009. Designing utterances for action: Verb repeat responses to assessments. Julkaisussa: Haakana, Markku, Laakso, Minna & Lindström, Jan (toim.) *Talk in Interaction. Comparative Dimensions*. Helsinki: Finnish Literature Society, 124–151.

Hakulinen, Auli & Sorjonen, Marja-Leena. 2012. Being equivocal: Affective responses left unspecified. Julkaisussa: Peräkylä, Anssi & Sorjonen, Marja-Leena (toim.) *Emotion in Interaction*. Oxford: Oxford University Press, 147–173.

Hall, Joan Kelly, Hellermann, John & Pekarek Doehler, Simona (toim.). 2011. *L2 Interactional Competence and Development*. Bristol: Multilingual Matters.

Harjunpää, Katariina. 2017. *Translatory Practices in Everyday Conversation: Bilingual Mediating in Finnish–Brazilian Portuguese Interaction*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN: 978-951-51-3227-7.

Harjunpää, Katariina, Mondada, Lorenza & Svinhufvud, Kimmo. 2018. The coordinated entry into service encounters in food shops: Managing interactional space, availability, and service during openings. *Research on Language and Social Interaction* 51:3, 271–291. Erikoisnumero *Entry and re-entry into interaction* (toim. Pillet-Shore, Danielle).

Harjunpää, Katariina, Mondada, Lorenza & Svinhufvud, Kimmo. 2019. Multimodaalinen litterointi keskustelunanalyysissa. *Puhe ja kieli* 39:3, 195–220.

Have, Paul ten. 2007. *Doing Conversation Analysis: A Practical Guide*. (2. p.). London: Sage.

Hayashi, Makoto. 2003. *Joint Utterance Construction in Japanese Conversation*. Amsterdam: John Benjamins.

Hayashi, Makoto, Raymond, Geoffrey, & Sidnell, Jack (toim.). 2013. *Conversational Repair and Human Understanding*. Cambridge: Cambridge University Press.

Hazel, Spencer. 2015. Identities at odds: embedded and implicit language policing in the internationalized workplace. *Language and Intercultural Communication* 15:1, 141–160.
Heath, Christian. 1986. *Body Movement and Speech in Medical Interaction.* Cambridge: Cambridge University Press.
Heinemann, Trine & Koivisto, Aino. 2016. Indicating a change-of-state in interaction: Cross-linguistic explorations. *Journal of Pragmatics* 104, 83–88.
Heinonen, Pilvi. 2017. *Evaluointi opettajan vuorovaikutustoimintana: Osallistujuuden ulottuvuudet luokkahuoneessa*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-3724-1.
Helasvuo, Marja-Liisa. 2004. Shared syntax: the grammar of co-constructions. *Journal of Pragmatics* 36, 1315–1336.
Helasvuo, Marja-Liisa. 2008. Minä ja muut: Puhujaviitteisyys ja kontekstuaalinen tulkinta. *Virittäjä* 112:2, 186–206.
Helasvuo, Marja-Liisa. 2009. Emergent grammar. Julkaisussa: Brisard, Frank, Östman, Jan-Ola & Verschueren, Jef (toim.) *Grammar, Meaning and Pragmatics. Handbook of Pragmatics Highlights* 5. Amsterdam: John Benjamins, 66–73.
Helasvuo, Marja-Liisa. 2014. Agreement or crystallization: Patterns of 1st and 2nd person subjects and verbs of cognition in Finnish conversational interaction. *Journal of Pragmatics* 63, 63–78.
Helasvuo, Marja-Liisa, Endo, Tomoko & Kärkkäinen, Elise (toim.). 2018. Units in responsive turns. Erikoisnumero. *Journal of Pragmatics* 123, 117–120.
Helasvuo, Marja-Liisa & Huumo, Tuomas. 2010. Mikä subjekti on? *Virittäjä* 114:2, 165–195.
Henricson, Sofie. 2013. *Svenska i finsk miljö: Interaktion, grammatik och flerspråkighet i samtal på svenska språköar i Finland*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-9167-4.
Hepburn, Alexa & Bolden, Galina. 2012. The conversation analytic approach to transcription. Julkaisussa: Sidnell, Jack & Stivers, Tanya (toim.) *The Handbook of Conversation Analysis*. New Jersey: Wiley-Blackwell, 57–76.
Heritage, John. 1996 [1984]. *Garfinkel ja etnometodologia*. Suomentaneet Ilkka Arminen, Outi Paloposki, Anssi Peräkylä, Sanna Vehviläinen & Soile Veijola. Helsinki: Gaudeamus.
Heritage, John. 2012a. Epistemics in action: Action formation and territories of knowledge. *Research on Language and Social Interaction* 45:1, 1–29.
Heritage, John. 2012b. The epistemic engine: Sequence organization and territories of knowledge. *Research on Language and Social Interaction* 45:1, 30–52.
Heritage, John & Clayman, Steven. 2010. *Talk in Action. Interactions, Identities, and Institutions*. Chichester: Wiley-Blackwell.
Heritage, John & Sorjonen, Marja-Leena. 1994. Constituting and maintaining activities across sequences: *and*-prefacing as a feature of question design. *Language in Society* 23, 1–29.
Hopper, Paul. 1987. Emergent grammar. *Berkeley Linguistic Society* 13, 139–157.
Hopper, Paul. 2011. Emergent grammar and temporality in interactional linguistics. Julkaisussa: Auer, Peter & Pfänder, Stefan (toim.) *Constructions: Emerging and Emergent*. Berlin: De Gruyter, 22–44.

Humă, Bogdana, Stokoe, Elizabeth & Sikveland, Rein Ove. 2019. Persuasive conduct: Alignment and resistance in prospecting ”cold” calls. *Journal of Language and Social Psychology* 38:1, 33–60.

ISK = Hakulinen, Auli, Vilkuna, Maria, Korhonen, Riitta, Koivisto, Vesa, Heinonen, Tarja Riitta & Alho, Irja. 2004. *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Karvonen, Ulla, Heinonen, Pilvi & Tainio, Liisa. 2018. Kosketuksen lukutaitoa. Konventionaalistuneet kosketukset oppilaiden välisessä vuorovaikutuksessa. *AFinLan vuosikirja* 76:1, 159–183. Saatavissa: https://doi.org/10.30661/afinlavk.69257.

Kaski, Timo, Niemi, Jarkko & Pullins, Ellen. 2018. Rapport building in authentic B2B sales interaction. *Industrial Marketing Management* 69, 235–252.

Keevallik, Leelo. 2018. What does embodied interaction tell us about grammar? *Research on Language and Social Interaction* 51:1, 1–21.

Keisanen, Tiina, Rauniomaa, Mirka & Siitonen, Pauliina. 2017. Transitions as sites of socialization in family interaction outdoors. *Learning, Culture and Social Interaction* 14, 24–37.

Klippi, Anu. 1996. *Conversation as an Achievement in Aphasics*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Koivisto, Aino. 2011. *Sanomattakin selvää?* Ja, mutta *ja* että *puheenvuoron lopussa*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-7323-6.

Koivisto, Aino. 2017. Uutta tietoa vai oivallus? Eräiden dialogipartikkeleiden tehtävistä. *Virittäjä* 121:4, 473–499.

Koivisto, Aino & Nykänen, Elise (toim.). 2013. *Dialogi kaunokirjallisuudessa*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Koivisto, Aino & Nykänen, Elise (toim.). 2016. Approaches to fictional dialogue. Erikoisnumero. *International Journal of Literary Linguistics* 5:2.

Korpela, Eveliina. 2007. *Oireista puhuminen lääkärin vastaanotolla: Keskustelunanalyyttinen tutkimus lääkärin kysymyksistä*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Koskela, Heidi & Piirainen-Marsh, Arja. 2002. Harvey Sacks ja keskusteluntutkimuksen perinne. Julkaisussa: Dufva, Hannele & Lähteenmäki, Mika (toim.) *Kielentutkimuksen klassikoita*. Soveltavan kielentutkimuksen teoriaa ja käytäntöä. Jyväskylä: Soveltavan kielentutkimuksen keskus, 257–280.

Kotilainen, Lari, Kurhila, Salla & Kalliokoski, Jyrki (toim.). 2019. *Kielenoppiminen luokan ulkopuolella*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kurhila, Salla. 2006. *Second Language Interaction*. Amsterdam: John Benjamins.

Kurhila, Salla & Kotilainen, Lari. 2017. Cooking, interaction and learning: The Finnish digital kitchen as a language-learning environment. Julkaisussa: Seedhouse, Paul (toim.) *Task-based Language Learning in a Real-world Digital Environment*. London: Bloomsbury Academic, 157–179.

Kärkkäinen, Elise & Keisanen, Tiina. 2012. Linguistic and embodied formats for making (concrete) offers. *Discourse Studies* 14:5, 587–611.

Kääntä, Leila. 2015. The multimodal organisation of teacher-led classroom interaction. Julkaisussa: Jenks, Christopher J. & Seedhouse, Paul (toim.) *International Perspectives on ELT Classroom Interaction*. London: Palgrave Macmillan, 6–83.

Laakso, Minna. 1997. *Self-initiated Repair by Fluent Aphasic Speakers in Conversation*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Laakso, Minna. 2005. Afaattisten henkilöiden vuorovaikutuksen keskustelunanalyyttinen tutkimus. *Puhe ja kieli* 25:2, 53–64.
Laakso, Minna. 2015. Collaborative participation in aphasic word searching: Comparison between significant others and speech and language therapists. *Aphasiology* 29:3, 269–290.
Labov, William. 1972. The study of language in its social context. Julkaisussa: Giglioli, Pier Paolo (toim.) *Language and Social Context*. London: Penguin, 283–307.
Labov, William & Fanshel, David. 1977. *Therapeutic Discourse*. New York (NY): Academic Press.
Lappalainen, Hanna & Raevaara, Liisa (toim.). 2009. *Kieli kioskilla. Tutkimuksia kioskiasioinnin rutiineista*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Lauranto, Yrjö. 2015. *Direktiivisyyden rajoja. Suomen kielen vaihtokauppasyntaksia*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-1180-7.
Laury, Ritva. 1997. *Demonstratives in Interaction*. Amsterdam: John Benjamins.
Laury, Ritva & Suzuki, Ryoko (toim.). 2011. *Subordination in Conversation: A Cross-linguistic Perspective*. Amsterdam: John Benjamins.
Lehtinen, Esa & Pälli, Pekka. 2016. Osallistuminen näkökulmana tekstien ja keskustelujen suhteeseen: Kehityskeskustelulomake osana kehityskeskustelun vuorovaikutusta ja organisatorisia käytänteitä. *Virittäjä* 120:2, 164–187.
Leskelä, Leealaura & Lindholm, Camilla (toim.). 2012. *Haavoittuva keskustelu. Keskustelunanalyyttisia tutkimuksia kielellisesti epäsymmetrisestä vuorovaikutuksesta*. Kehitysvammaliiton tutkimuksia nro 6. Helsinki: Kehitysvammaliitto ry.
Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
Lilja, Niina & Tapaninen, Terhi. 2019. Suomen kielen käyttämisen ja oppimisen mahdollisuudet ammatillisen oppilaitoksen rakennusalan vuorovaikutustilanteissa. *Puhe ja kieli* 39:1, 69–98.
Lindfors, Pirjo & Raevaara, Liisa. 2005. Discussing patients' drinking and eating habits in medical and homeopathic consultation. *Communication and Medicine* 2:2, 137–149.
Lindholm, Camilla & Leskelä, Leealaura. 2018. Vuorovaikutuksen epäsymmetria. *Memo* 1, 16–17.
Lindström, Jan. 2006. Grammar in the service of interaction: Exploring turn organization in Swedish. *Research on Language and Social Interaction* 39:1, 81–117.
Lindström, Jan, Maschler, Yael & Pekarek Doehler, Simona. 2016. A cross-linguistic perspective on grammar and negative epistemics in talk-in-interaction. *Journal of Pragmatics* 106, 72–79.
Margutti, Piera, Tainio, Liisa, Drew, Paul & Traverso, Véronique (toim.). 2018. Inviting in telephone calls: a cross-linguistic study of social actions in interaction. Erikoisnumero. *Journal of Pragmatics* 125, 52–199.
Maynard, Douglas W. 2012. Everyone and no one to turn to: Intellectual roots and contexts for conversation analysis. Julkaisussa: Sidnell, Jack & Stivers, Tanya (toim.) *The Handbook of Conversation Analysis*. Chichester: Wiley-Blackwell, 11–31.

Meredith, Joanne. 2019. Conversation analysis and online interaction. *Research on Language and Social Interaction* 52:3, 241–256.
Merke, Saija. 2016. *Opiskelijoiden aloittamat kysymyssekvenssit suomi vieraana kielenä -oppitunnin voimavarana*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-2193-6.
Mikkola, Piia & Lehtinen, Esa. 2019. Drawing conclusions about what co-participants know: Knowledge-probing question–answer sequences in new employee orientation lectures. *Discourse and Communication* 13:5, 516–538.
Mondada, Lorenza. 2012a. The conversation analytic approach to data collection. Julkaisussa: Sidnell, Jack & Stivers, Tanya (toim.) *The Handbook of Conversation Analysis*. Chichester: Wiley-Blackwell, 32–56.
Mondada, Lorenza. 2012b. The dynamics of embodied participation and language choice in multilingual meetings. *Language in Society* 41:2, 213–235.
Mondada, Lorenza 2016. Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics* 20:3, 336–366.
Mondada, Lorenza. 2018. Greetings as a device to find out and establish the language of service encounters in multilingual settings. *Journal of Pragmatics* 126, 10–28.
Mondada, Lorenza & Svinhufvud, Kimmo. 2016. Writing-in-interaction. Studying writing as a multimodal phenomenon in social interaction. *Language and Dialogue* 6:1, 1–53.
Monzoni, Chiara & Laury, Ritva. 2015. Making referents accessible in multi-party interaction. *Journal of Estonian and Finno-Ugric Linguistics* 6:2, 43–62.
Nevile, Maurice. 2015. The embodied turn in research on language and social interaction. *Research on Language and Social Interaction* 48:2, 121–151.
Nevile, Maurice, Haddington, Pentti, Heinemann, Trine & Rauniomaa, Mirka. 2014. *Interacting with Objects. Language, Materiality, and Social Activity*. Amsterdam: John Benjamins.
Niemi, Jarkko. 2010. Myönnyttelymuotti – Erimielisyyttä enteilevä samanmielisyyden konstruktio. *Virittäjä* 114:2, 196–222.
Niemi, Jarkko. 2013. Tiedon rajat ja vuorovaikutus. Toteamukseen tai vaihtoehtokysymykseen vastaavat voi olla -rakenteet. *Virittäjä* 117:2, 201–241.
Niemi, Jarkko. 2015. *Myönnyttelyn käytänteitä. Erimielisyys ja yhteisymmärryksen rakentaminen vuorovaikutuksessa*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-0603-2.
Niemi, Jarkko & Hirvonen, Linda. 2019. Money talks: Customer-initiated price negotiation in business-to-business sales interaction. *Discourse & Communication* 13:1, 95–118.
Nilep, Chad. 2006. "Code switching" in sociocultural linguistics. *Colorado Research in Linguistics* 19:1, 1–22.
Nissi, Riikka & Lehtinen, Esa. 2016. Negotiation of expertise and multifunctionality: PowerPoint presentations as interactional activity types in workplace meetings. *Language & Communication* 48, 1–17.
Norrby, Catrin, Wide, Camilla, Nilsson, Jenny & Lindström, Jan. 2018. Positioning through address practice in Finland-Swedish and Sweden-Swedish service encounters. Julkaisussa: Beeching, Kate, Ghezzi, Chiara & Molinelli, Piera (toim.) *Positioning the Self and Others: Linguistic Perspectives*. Amsterdam: John Benjamins, 19–49.

Ochs, Elinor, Schegloff, Emanuel A. & Thompson, Sandra A. 1996. *Interaction and Grammar.* Cambridge: Cambridge University Press.

Ogden, Richard. 2001. Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association* 43:3, 299–320.

Ogden, Richard. 2013. Clicks and percussives in English conversation. *Journal of the International Phonetic Association* 43:3, 299–320.

Ogden, Richard, Hakulinen, Auli & Tainio, Liisa. 2004. Indexing 'no news' with stylization in Finnish. Julkaisussa: Couper-Kuhlen, Elizabeth & Ford, Cecilia (toim.) *Sound Patterns in Interaction: Cross-linguistic Studies from Conversation*. Amsterdam: John Benjamins, 299–334.

Ogden, Richard & Routarinne, Sara. 2005. The communicative functions of final rises in Finnish intonation. *Phonetica* 62:2–4, 160–175.

Ono, Tsuyoshi, Laury, Ritva & Suzuki, Ryoko. 2019. On the notion of unit in the study of human languages. Erikoisnumero. *Studies in Language* 43:2, 245–253.

Paananen, Jenny. 2017. Kuinka lääkärit korjaavat kysymyksiään? Kysymysten uudelleenmuotoilu monikulttuurisilla lääkärin vastaanotoilla. *Virittäjä* 120:4, 552–579.

Paananen, Jenny. 2019. *Yhteisymmärryksen rakentaminen monikulttuurisilla lääkärin vastaanotoilla*. Väitöskirja. Turku: Turun yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-29-7557-0.

Pajo, Kati. 2013. *Joint Multimodal Management of Hearing Impairment in Conversations at Home – Implications for Communication Therapy*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-9413-2.

Pekarek Doehler, Simona. 2011. Emergent grammar for all practical purposes: The on-line formatting of left and right dislocations in French conversation. Julkaisussa: Auer, Peter & Pfänder, Stefan (toim.) *Constructions: Emerging and Emergent*. Berlin: De Gruyter, 45–87.

Pekarek Doehler, Simona, Wagner, Johannes, González-Martínez, Esther (toim.). 2018. *Longitudinal Studies on the Organization of Social Interaction*. London: Palgrave.

Peräkylä, Anssi. 1997. Validity and reliability in research based on tapes and transcripts. Julkaisussa: Silverman, David (toim.) *Qualitative Analysis: Issues of Theory and Method*. London: Sage, 201–220.

Plejert, Charlotta, Lindholm, Camilla & Schrauf, Robert W. (toim.). 2017. *Multilingual Interaction and Dementia*. Bristol: Multilingual matters.

Pomerantz, Anita & Heritage, John. 2013. Preference. Julkaisussa: Sidnell, Jack & Stivers, Tanya (toim.) *The Handbook of Conversation Analysis*. Chichester: Wiley-Blackwell, 210–228.

Raevaara, Liisa. 2004. ”Mitäs me sovittais”. *-s*-partikkelin sisältävien hakukysymysten tehtävistä. *Virittäjä* 108:4, 531–558.

Raevaara, Liisa. 2009. ”Nyt täytyy ostaa kun tuli vähä rahaa tosta monivedosta”. Asiakkaan ja myyjän kootut selitykset. Julkaisussa: Lappalainen, Hanna & Raevaara, Liisa (toim.) *Kieli kioskilla. Tutkimuksia kioskiasioinnin rutiineista*. Helsinki: Suomalaisen Kirjallisuuden Seura, 231–262.

Raevaara, Liisa. 2011. Accounts at convenience stores: Doing dispreference and small talk. *Journal of Pragmatics* 43:2, 556–571.

Raevaara, Liisa. 2016. Toimintajaksojen rakenteet. Julkaisussa: Stevanovic, Melisa & Lindholm, Camilla (toim.) *Keskustelunanalyysi. Kuinka tutkia sosiaalista toimintaa ja vuorovaikutusta*. Tampere: Vastapaino, 143–161.

Raevaara, Liisa, Ruusuvuori, Johanna & Haakana, Markku. 2001. Institutionaalinen vuorovaikutus ja sen tutkiminen. Julkaisussa: Ruusuvuori, Johanna, Haakana, Markku & Raevaara, Liisa (toim.) *Institutionaalinen vuorovaikutus. Keskustelunanalyyttisia tutkimuksia*. Helsinki: Suomalaisen Kirjallisuuden Seura, 11–38.

Raevaara, Liisa & Sorjonen, Marja-Leena. 2006. Vuorovaikutuksen osanottajien toiminta ja genre. Keskustelunanalyysin näkökulma. Julkaisussa: Mäntynen, Anne, Shore, Susanna & Solin, Anna (toim.) *Genre – tekstilaji*. Helsinki: Suomalaisen Kirjallisuuden Seura, 122–150.

Raevaara, Liisa, Sorjonen, Marja-Leena & Lappalainen, Hanna. 2013. Vuorovaikutus Kelan puhelinpalvelussa. Kelan työpapereita 46. Saatavissa: http://hdl.handle.net/10138/39251.

Rossi, Giovanni. 2015. *The Request System in Italian Interaction*. Väitöskirja. Nijmegen: Radboud University. Saatavissa: http://hdl.handle.net/2066/141385.

Ruusuvuori, Johanna, Haakana, Markku & Raevaara, Liisa. 2001. *Institutionaalinen vuorovaikutus. Keskustelunanalyyttisia tutkimuksia*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Sacks, Harvey. 1992. *Lectures on Conversation*. Vol. I–II. Toimittanut Gail Jefferson. Oxford: Blacwell.

Sacks, Harvey, Schegloff, Emanuel A. & Jefferson, Gail. 2016 [1974]. Keskustelun vuorottelujäsennyksen perussystematiikka. Suomentaneet Marja Etelämäki, Auli Hakulinen, Ilkka Arminen, Elizabeth Couper-Kuhlen, Mia Halonen, Aino Koivisto, Salla Kurhila, Ritva Laury, Esa Lehtinen, Camilla Lindholm, Jarkko Niemi, Anssi Peräkylä, Johanna Ruusuvuori, Liisa Raevaara, Sara Routarinne, Melisa Stevanovic, Kimmo Svinhufvud, Liisa Tainio & Liisa Voutilainen. *Virittäjä* 120:3 [verkkoliite].

Savijärvi, Marjo. 2011. *Yhteisestä toiminnasta yhteiseen kieleen. Keskustelunanalyyttinen tutkimus toisen kielen oppimisesta kielikylpypäiväkodin arkitilanteissa*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-7309-0.

Schegloff, Emanuel A. 1968. Sequencing in conversational openings. *American Anthropologist* 70 6, 1075–1095.

Schegloff, Emanuel A. 1984. Some sources of misunderstanding in talk-in-interaction. *Linguistics* 25, 201–218.

Schegloff, Emanuel A. 1992. Introduction. Julkaisussa: Sacks, Harvey *Lectures on conversation I*. Toimittanut Gail Jefferson. Oxford: Blackwell, ix–lii.

Schegloff, Emanuel A. 1996. Turn organization: one intersection of grammar and interaction. Julkaisussa: Ochs, Elinor, Schegloff, Emanuel A. & Thompson, Sandra A. (toim.) *Interaction and Grammar*. Cambridge: Cambridge University Press, 52–133.

Schegloff, Emanuel A. 2007. *Sequence Organization. A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.

Schegloff, Emanuel A., Jefferson, Gail & Sacks, Harvey. 1977. The preference for self-correction in the organization of repair in conversation. *Language* 53:2, 361–381.

Schegloff, Emanuel A., Ochs, Elinor & Thompson, Sandra A. 1996. Introduction. Julkaisussa: Ochs, Elinor, Schegloff, Emanuel A. & Thompson, Sandra A. (toim.) *Interaction and Grammar.* Cambridge: Cambridge University Press, 1–51.

Schegloff, Emanuel A. & Sacks, Harvey. 1973. Opening up closings. *Semiotica* 7, 289–327.
Selting, Margret, Auer, Peter, Barden, Birgit, Bergmann, Jörg, Couper-Kuhlen, Elizabeth, Günthner, Susanne, Meier, Christoph, Quasthoff, Uta, Schlobinski, Peter, Uhmann, Susanne. 2011. A transcription system for conversation analysis. Saksankielinen alkuperäisartikkeli 1998. Kääntäneet ja sovittaneet englantiin Elizabeth Couper-Kuhlen & Dagmar Barth-Weingarten. *Gesprächsforschung. Online-Zeitschrift zur verbalen Interaktion* 12, 1–51.
Seppänen, Eeva-Leena. 1997a. Suomen perfektin merkityksestä keskusteluaineiston valossa. *Virittäjä* 101, 2–26.
Seppänen, Eeva-Leena. 1997b. Vuorovaikutus paperilla. Julkaisussa: Tainio, Liisa (toim.) *Keskustelunanalyysin perusteet*. Tampere: Vastapaino, 18–31.
Seppänen, Eeva-Leena. 1997c. Osallistumiskehikko. Julkaisussa: Tainio, Liisa (toim.) *Keskustelunanalyysin perusteet*. Tampere: Vastapaino, 156–176.
Seppänen, Eeva-Leena. 1998. *Läsnäolon pronominit*: Tämä, tuo, se *ja* hän *viittaamassa keskustelun osallistujaan*. Väitöskirja. Helsinki: Suomalaisen Kirjallisuuden Seura.
Seppänen, Eeva-Leena & Herlin, Ilona. 2009. Kuvauksista seurauksiin: Kaksiulotteinen konstruktio *niin* A *että*. *Virittäjä* 113:2, 213–245.
Sidnell, Jack. 2009a. *Conversation Analysis. An Introduction.* Chichester: Wiley-Blacwell.
Sidnell, Jack. 2009b. *Conversational Analysis. Comparative Perspectives.* Cambridge: Cambridge University Press.
Sidnell, Jack. 2012. Basic conversation analytic methods. Julkaisussa: Sidnell, Jack & Stivers, Tanya (toim.) *The Handbook of Conversation Analysis.* Chichester: Wiley-Blackwell, 77–99.
Sidnell, Jack & Stivers, Tanya (toim.). 2012. *The Handbook of Conversation Analysis.* Chichester: Wiley-Blackwell.
Siitonen, Pauliina, Rauniomaa, Mirka & Keisanen, Tiina. 2019. ”Kato. Hulluna puolukoita.” *Kato* vuorovaikutuksen resurssina luontoilussa. *Virittäjä* 123:4, 518–549.
Sinclair, John & Coulthard, Malcolm. 1975. *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*. Oxford: Oxford University Press.
Sorjonen, Marja-Leena. 1997. Korjausjäsennys. Julkaisussa: Tainio, Liisa (toim.) *Keskustelunanalyysin perusteet*. Tampere: Vastapaino, 112–138.
Sorjonen, Marja-Leena. 2001a. *Responding in Conversation. A Study of Response Particles in Finnish*. Amsterdam: John Benjamins.
Sorjonen, Marja-Leena. 2001b. Simple answers to polar questions: The case of Finnish. Julkaisussa: Selting, Margret & Couper-Kuhlen, Elizabeth (toim.) *Studies in Interactional Linguistics*. Amsterdam: John Benjamins, 405–429.
Sorjonen, Marja-Leena & Hakulinen, Auli. 2009. Alternative responses to assessments. Julkaisussa: Sidnell, Jack (toim.) *Conversation Analysis. Comparative Perspectives.* Cambridge: Cambridge University Press, 280–300.
Sorjonen, Marja-Leena, Peräkylä, Anssi & Eskola, Kari (toim). 2001. *Keskustelu lääkärin vastaanotolla*. Tampere: Vastapaino.
Sorjonen, Marja-Leena & Raevaara, Liisa. 2006. Asiointitilanteiden tutkimisesta. Tarkasteltavana Kelan vuorovaikutus. Julkaisussa: Sorjonen, Marja-Leena & Raevaara, Liisa (toim.) *Arjen asiointia. Keskusteluja Kelan tiskin äärellä*. Helsinki: Suomalaisen Kirjallisuuden Seura, 11–34.

Sorjonen, Marja-Leena & Raevaara, Liisa. 2014. On the grammatical form of requests at the convenience store: Requesting as embodied action. Julkaisussa: Drew, Paul & Couper-Kuhlen, Elizabeth (toim.) *Requesting in Social Interaction*. Amsterdam: John Benjamins, 243–268.

Sorjonen, Marja-Leena, Raevaara, Liisa & Couper-Kuhlen, Elizabeth (toim.). 2017. *Imperative Turns at Talk*. Amsterdam: John Benjamins.

Sorjonen, Marja-Leena, Rouhikoski, Anu & Lehtonen, Heini (toim.). 2015. *Helsingissä puhuttavat suomet: Kielen indeksisyys ja sosiaaliset identiteetit*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Sovijärvi, Antti & Peltola, Reijo. 1977. *Suomalais-ugrilainen tarkekirjoitus*. Helsingin yliopiston fonetiikan laitoksen julkaisuja 9. Helsinki: Helsingin yliopisto.

Stevanovic, Melisa & Lindholm, Camilla (toim.). 2016. *Keskustelunanalyysi. Kuinka tutkia sosiaalista toimintaa ja vuorovaikutusta*. Tampere: Vastapaino.

Stevanovic, Melisa & Weiste, Elina. 2017. Conversation analytic data session as a pedagogical institution. *Learning, Culture, and Social Interaction* 15, 1–17.

Stommel, Wyke, Paulus, Trena & Atkins, David. 2017. "Here's the link": Hyperlinking in service-focused chat interaction. *Journal of Pragmatics* 115, 56–67.

Streeck, Jürgen. 2018. Grammaticalization and bodily action: Do they go together? *Research on Language and Social Interaction* 51:1, 26–32.

Svinhufvud, Kimmo. 2018. Waiting for the customer: Multimodal analysis of waiting in service encounters. *Journal of Pragmatics* 129, 48–75.

Tainio, Liisa. 1997. Alkusanat. Julkaisussa: Tainio, Liisa (toim.) *Keskustelunanalyysin perusteet*. Tampere: Vastapaino, 9–11.

Thompson, Sandra & Couper-Kuhlen, Elizabeth. Tulossa. English *why don't you X* as a formulaic expression.

Thompson, Sandra, Fox, Barbara & Couper-Kuhlen, Elizabeth. 2015. *Grammar in Everyday Talk: Building Responsive Actions*. Cambridge: Cambridge University Press.

Tieteen termipankki. 2019. [verkkoaineisto]. [viitattu 3.4.2018]. http://www.tieteentermipankki.fi.

Tiitinen, Sanni. 2015. *Supporting Parenthood in Interaction. A Conversation Analytic Study of Maternity and Child Health Clinics*. Väitöskirja. Acta Universitatis Tamperensis 2009. Tampere: Tampere University Press. Saatavissa: http://urn.fi/URN:ISBN:978-951-44-9673-8.

Traverso, Véronique. 2012. Ad hoc-interpreting in multilingual workplaces: Who translates for whom? Julkaisussa: Baraldi, Claudio & Gavioli, Laura (toim.) *Coordinating Participation in Dialogue Interpreting*. Amsterdam: John Benjamins, 149–176.

Vatanen, Anna. 2014. *Responding in Overlap. Agency, Epistemicity and Social Action in Conversation*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-51-0280-5.

Vatanen, Anna. 2016. Keskustelunanalyyttinen tutkimusprosessi. Julkaisussa: Stevanovic, Melisa & Lindholm, Camilla (toim.) *Keskustelunanalyysi. Kuinka tutkia sosiaalista toimintaa ja vuorovaikutusta*. Tampere: Vastapaino, 312–330.

Vepsäläinen, Heidi. 2019. *Suomen "no"-partikkeli ja kysymyksiin vastaaminen keskustelussa*. Väitöskirja. Helsinki: Helsingin yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-7309-0.

Vilkuna, Maria. 1992 [1980]. *Referenssi ja määräisyys suomenkielisten tekstien tulkinnassa*. Helsinki: Suomalaisen Kirjallisuuden Seura.
Weiste, Elina & Voutilainen, Liisa. 2016. Empatian ja erimielisyyden ilmaiseminen psykoterapiassa: vuorovaikutustutkimuksen näkökulma. Julkaisussa: Turpeinen, Päivi & Nurmento, Riikka (toim.) *Terapiasuhteessa olemisen taito.* Helsinki: Edita, 212–244.
Whalen, Jack, Zimmerman, Don & Whalen, Marilyn. 1988. When words fail: a single case analysis. *Social Problems* 35:4, 335–362.
Zinken, Jörg. 2016. *Requesting Responsibility. The Morality of Grammar in Polish and English Family Interaction*. Oxford: Oxford University Press.

# Liitteet

## Liite 1

### KESKUSTELUNANALYYTTISET PUHEEN LITTEROINTIMERKIT

Sävelkulku ja sävelkorkeus:

| | |
|---|---|
| . | laskeva sävelkulku |
| ; | lievästi laskeva sävelkulku |
| , | tasainen sävelkulku |
| ? | nouseva sävelkulku |
| ↑ | ympäröivää puhetta korkeammalta aloitettu tai sanottu jakso |
| ↓ | ympäröivää puhetta matalammalta aloitettu tai sanottu jakso |

Painotus ja puhenopeus:

| | |
|---|---|
| _ | painotettu sana tai sanan osa on alleviivattu (esim. *se lomake*) |
| ° | ympäröivää puhetta hiljaisemmin sanottu jakso |
| > < | ympäröivää puhetta nopeammin sanottu jakso |
| < > | ympäröivää puhetta hitaammin sanottu jakso |

Sanojen kesto:

| | |
|---|---|
| - | kesken jäänyt sana (esim. *se lap- eiku lomake)* |
| < | sanan loppu tavallista lyhyemmin sanottu (esim. *se lisä vai mikä*<) |
| : | äänteen venytys (esim. *se ko::rkea talo*) |
| _ | kaksi sanaa on lausuttu yhteen; legatoääntämys (esim. *täss_on tää*) |

Muita esitystapaa koskevia seikkoja:

| | |
|---|---|
| £ | hymyillen sanottu jakso |
| @ | äänen laadun muutos |
| (h) | naurua (tai hengitystä) sanan sisällä |
| .joo | piste sanan edessä: sana on lausuttu sisään hengittäen |
| .hhh | piste h:n edessä: sisään hengitys (yksi h on sekunnin kymmenesosa) |
| hhh | uloshengitys |

Tauot, päällekkäispuhunta:

| | |
|---|---|
| (0.7) | tauko ja sen kesto sekunteina |
| (.) | mikrotauko (alle 0.2 sekuntia) |
| (,) | mikrotaukoa pidempi mittaamaton tauko |
| = | tauotta toisiinsa liittyvät puhunnokset |
| [ | päällekkäispuhunnan alku |
| ] | päällekkäispuhunnan loppu |

Epäselvät kohdat ja litteroijan kommentit:

| | |
|---|---|
| (kas) | sulkeissa epäselvästi kuultu jakso |
| (-) | jakso, josta ei ole saatu selvää |

## Liite 2

Multimodaaliset litterointimerkit (täydempi versio ja ohjeistus saatavilla osoitteessa https://www.lorenzamondada.net/multimodal-transcription).

| | |
|---|---|
| * * | Kehollinen liike merkitään kahden |
| + + | samanlaisen symbolin väliin (erilainen symboli jokaista osallistujaa varten) ja kohdennetaan vastaavien puhejaksojen/taukojen kanssa. |
| *---> | Merkitty toiminta jatkuu toisella rivillä |
| ---->* | uuteen samanlaiseen symboliin saakka. |
| >> | Toiminta on alkanut ennen aineistokatkelman alkua. |
| --->> | Toiminta jatkuu aineistokatkelman päätyttyä. |
| ..... | Valmisteluvaihe. |
| ---- | Liikkeen huippuvaiheen/täyden laajuuden saavuttaminen ja ylläpito. |
| ,,,,, | Vetäytymisvaihe. |
| asi/myy | Liikkeen tuottajan tunniste merkitään, jos eri kuin senhetkinen puhuja. |
| Kuva | Kuvapysäytyksen tarkka ajoitus merkitään |
| # | meneillään olevaan vuoroon ristikkomerkillä. |

# Luku 23 Käännöstiede ja sen menetelmät

*Liisa Tiittula*
https://orcid.org/0000-0002-4688-0854

*Leena Kolehmainen*
https://orcid.org/0000-0003-2746-7201

**Mikä?**

*Käännöstiede* on termi, joka kattaa sekä kääntämisen että tulkkauksen tutkimuksen. Tutkimuksessa, jota tehdään sekä historiallisesta että nykyaikaan keskittyvästä näkökulmasta, käytetään erilaisia aineistoja ja menetelmiä tutkittaessa esimerkiksi seuraavia alueita:

- kääntämisen ja tulkkauksen tuotteita (käännöksiä ja tulkkeita),
- käännös- ja tulkkausprosessia,
- käännösten ja tulkkeiden vastaanottoa ja käyttöä (lukijoita, kuulijoita jne.),
- kääntäjiä, tulkkeja, kielentarkastajia, toimeksiantajia, kustantajia jne. (toimijoita),
- kääntämisen ja tulkkauksen toiminta- ja esiintymisympäristöjä,
- kääntäjien ja tulkkien tiedonhankintaa ja apuvälineitä sekä
- kääntäjien ja tulkkien koulutusta.

**Katso myös:**

Luku 5 Lingvistinen etnografia toimintayhteisöissä

Luku 8 Käsiteanalyysi kielentutkimuksessa

Luku 9 Korpusaineistot

Luku 10 Määrällinen korpuslingvistiikka

# 1. Johdanto

Roman Jakobson (2000 [1959]) jakoi kääntämisen kieltenväliseen, kielensisäiseen ja intersemioottiseen kääntämiseen. Näistä ensiksi mainittu ja sen tutkimus muodostavat käännöstieteellisen tutkimuksen perinteisen ytimen. Yhä useammat tutkijat tarkastelevat kuitenkin myös kielensisäistä (mukaelmat, siirtymä kirjoitetusta moodista suulliseen ja päinvastoin) ja intersemioottista (siirtymä merkkijärjestelmästä toiseen) kääntämistä, sillä monet käännösviestinnän osa-alueet sisältävät komponentteja useammasta Jakobsonin erottamasta luokasta (esim. kirjoitustulkkaus, kuvailutulkkaus ja ohjelmatekstitys).

Kääntäminen ja sen tutkimus on ollut tapana jakaa osa-alueisiin tekstilajeittain kirjallisuuden kääntämiseen (kauno- ja tietokirjallisuus), asiatekstien kääntämiseen (lait, sopimukset, todistukset, esitteet jne.) ja av-kääntämiseen (audiovisuaalinen kääntäminen: tv-ohjelmien, elokuvien, oopperan jne. tekstitys, selostusteksti, ohjelmatekstitys, dubbaus, verkkosivujen ja pelien kääntäminen). Tulkkausta ja sen tutkimusta taas voidaan lähestyä tulkkausmenetelmästä (simultaani-, konsekutiivi-, kuiskaustulkkaus) tai tulkkaustilanteesta käsin (oikeus-, asioimis-, konferenssi-, neuvottelutulkkaus), teknologiset mahdollisuudet huomioiden (lähi- ja etätulkkaus), kielimuodon näkökulmasta (puhuttujen kielten vs. viittomakielen tulkkaus) tai erityiskohderyhmä ja aistirajoite huomioiden (kuvailutulkkaus, kirjoitustulkkaus,

puhevammaistulkkaus, taktiilitulkkaus). Mahdollisesti tuorein jaottelu koskee toimijoita: maallikko-, fani-, vapaaehtois- ja joukkoistetun kääntämisen ja tulkkauksen tutkimus on ammattikääntäjien ja -tulkkien toiminnan tutkimuksen rinnalle kehittynyt uusi tutkimussuuntaus. Käännöstieteen alaan luetaan usein myös lähikäsitteet *tekninen viestintä* ja *lokalisointi*: Ensimmäisellä tarkoitetaan työtehtäviä, joissa henkilö vastaa erilaisten teknisten tekstien ja dokumenttien, kuten käyttöohjeiden tai huolto- ja asennusoppaiden, kirjoittamisesta ja päivittämisestä. Jälkimmäinen taas viittaa tuotteiden, kuten tietokoneohjelmien ja -pelien, sovittamiseen eri kieli- ja kulttuuriympäristöjen käyttäjille sopiviksi.

Joskus käännöstieteen rinnalle nostetaan omana suuntauksenaan tulkkaukseen keskittyvä tulkkaustiede. Tässä artikkelissa käytämme yläkäsitteenä *käännöstiedettä* ja tarkastelemme sekä kääntämistä että tulkkausta. Koska käännöstiede on tutkimuskohteiltaan monialainen, tutkijat käyttävät samoja menetelmiä kuin lähitieteenaloilla käytetään, erityisesti kielitieteessä (esim. korpustutkimus, vertaileva tekstintutkimus), kirjallisuustieteessä ja yhteiskuntatieteissä (esim. kysely, haastattelu, havainnointi). Yhtymäkohtia on lukuisiin muihinkin tieteenaloihin, esimerkiksi kulttuurintutkimukseen, historiantutkimukseen, tietojenkäsittelytieteeseen ja kognitiotieteeseen.

Seuraavassa luodaan katsaus käännöstieteelliseen tutkimukseen, siihen, millaisin aineistoin ja analyysimenetelmin ja mihin tarkoituksiin kieltä, tekstejä ja vuorovaikutusta analysoidaan käännöstieteessä. Havainnollistamme tutkijoiden hyödyntämiä aineistoja ja aineistonkeruumenetelmiä sekä analyyseissa käyttämiä metodeja. Käännöstieteelliset tutkimukset ovat yhä useammin monimetodisia ja moniaineistoisia: tutkijat lähestyvät tutkimuskysymystään eri näkökulmista ja valottavat sitä eri aineistoin ja metodein, ja monimenetelmäisellä otteella voidaan vastata uudenlaisiin tutkimuskysymyksiin.

Keskitymme kieleen, kirjoitettuihin ja puhuttuihin teksteihin, niiden tuottamisen ja vastaanottamisen prosesseihin sekä niiden käännöstieteelliseen analyysiin ja toimijoihin. Tästä syystä esimerkiksi kääntäjien tiedonhankintaa ja apuvälineitä sekä koulutusta koskeva käännöstieteellinen tutkimus rajautuu artikkelin ulkopuolelle. Lähitutkimussuuntauksista tarkastelematta jäävät terminologinen

(ks. Käsiteanalyysi kielentutkimuksessa tässä kirjassa) ja viittomakielen tulkkauksen tutkimus, joista molemmat ansaitsisivat kokonaan omat artikkelinsa. Myös konekääntäminen ja käännösteknologia omana tutkimuskohteenaan jää pois.

Luomme ensin lyhyen katsauksen käännöstieteen historiaan (luku 2). Sitä seuraa kuvaus käännöstieteen tutkimuskohteista, metodeista ja aineistoista (luku 3): esittelemme ensin historiallisesti perinteikkäimmän tutkimuskohteen eli käännettyjen tekstien ja tulkkeiden tutkimuksen (luvut 3.1. ja 3.2.), jonka jälkeen siirrymme kääntämisen ja tulkkauksen prosessien (3.3.) ja siihen läheisesti liittyvään toimijoiden, toimintaympäristön ja vastaanoton tutkimukseen (3.4.). Viimeinen luku 4 kokoaa keskeisiä tendenssejä yhteen.

## 2. Käännöstieteen historiaa ja taustaa

**Kielikäsitys**

Käännöstieteessä käsitys kielestä vaihtelee niin tutkimussuunnittain kuin historiallisesti. Kieli nähdään usein välineenä: yhteinen kieli mahdollistaa vuorovaikutuksen eri osapuolten välille, yhteisen kielen puuttuminen taas muodostaa esteen viestinnälle. Kääntäjät ja tulkit luovat sillan osapuolten välille, kun erilaiset kieliresurssit ovat vuorovaikutuksen esteenä tai kun aistivamma tai muut toimintarajoitteet vaikeuttavat viestinnän saavutettavuutta. He mahdollistavat kielellisten oikeuksien toteutumisen tilanteissa, joissa oman kielen käyttö taataan esimerkiksi erikielisten viranomaisten kanssa asioitaessa.

Käännöstieteessä kielikäsitykseen kuuluu käsitys kielten välisistä suhteista eli toisenkielisen käännöksen tai tulkkeen tuottamisen ehdoista, siitä, miten yhden kielen viestin pohjalta tuotetaan toisella kielellä uusi viesti uudelle vastaanottajalle ja uuteen tarkoitukseen. Käännöstieteen historiassa, erityisesti konekääntämisen alkuaikoina, kieltenvälisiä suhteita pidettiin varsin suoraviivaisina vastaavuuksina ja kääntäminen näyttäytyi

mekaanisena toimintana. Nykyään painottuu näkemys kieltenvälisten suhteiden kompleksisuudesta. Kieltä tarkastellaankin usein suhteessa kulttuuriin, puhujayhteisöihin, erilaisiin yleisöihin ja vastaanottajaprofiileihin, ideologisiin kysymyksiin, kuten valtaan, sekä vuorovaikutuksen esiintymiskontekstiin: Uuden erikielisen viestin funktio ja kontekstin tai toimintaympäristön sanelemat reunaehdot muodostavat keskeisiä kääntämiseen ja tulkkaukseen vaikuttavia asioita. Lisäksi kieli toimii osana multimodaalista kokonaisuutta. Kieli on käännöstieteessä keskeinen tiedon vaihtoon ja liikkumiseen liittyvä väline (engl. *knowledge transfer*). Se on myös esteettinen ilmiö, joka mahdollistaa taide-elämyksen.

Kirjallisen kääntämisen ja suullisen tulkkauksen tutkimuksen perinteet ovat varsin erilaiset, minkä vuoksi niitä käsitellään seuraavassa ensin erikseen. Molemmissa kyse on kuitenkin pohjimmiltaan kielten- ja kulttuurienvälisen viestinnän välittämisestä. Myöskään raja kirjallisen ja suullisen välillä ei ole aina selvä. Esimerkiksi prima vista -tulkkauksessa tehdään kirjoitetusta lähdetekstistä valmistelematon suullinen käännös, tekstityksessä puolestaan käännetään puhetta kirjoitettuun muotoon.

## 2.1. Kääntämisen tutkimus

Kääntäminen toimintana on ikivanhaa, ja kääntämisen kysymyksiä on tarkasteltu kauan ennen kuin voi puhua varsinaisesta tutkimuksesta. Kääntämisen perinteisiä kysymyksiä ovat lähde- ja kohdetekstin suhde ja niiden vastaavuus sekä kääntämisen uskollisuus vs. vapaus (vrt. esim. Chesterman 2016 [1997]). Näiden kysymysten merkitystä länsimaisessa käännöstutkimuksessa ja kääntämistä koskevassa keskustelussa on selitetty uskonnollisten tekstien keskeisyydellä kääntämisen historiassa (esim. Prunč 2007). Toinen keskeinen alue on ollut kaunokirjallisuuden kääntäminen, jonka kysymyksiä on jo satojen vuosien ajan pohdittu.

Käännöksiä on tutkittu monilla tieteenaloilla: teologiassa on tutkittu Raamatun ja muiden pyhien tekstien käännöksiä, vertailevassa kirjallisuuden tutkimuksessa on tarkasteltu kirjallisuuden käännöksiä, ja

kielitieteessä erityisesti kontrastiivinen tutkimus on ollut kiinnostunut käännöksistä. Käännöstiede omana tieteenalana on varsin nuori; sen katsotaan syntyneen 1900-luvun loppupuoliskolla, jolloin se irtaantui keskeisistä emotieteistä, kieli- ja kirjallisuustieteistä (käännöstieteen historiasta ks. esim. Munday 2012; Prunč 2007).

Kielitieteen piirissä syntyi niin sanottu lingvistinen käännöstiede 1960-luvulla (mm. Catford 1965). Taustalla oli usko konekääntämiseen ja pyrkimys luoda kieliparien välille käännöksiä tuottavia sääntöjä, käännöskielioppeja. Tarkastelu kohdistui lähinnä sanatasolle, eikä lausetta suurempia kokonaisuuksia juuri huomioitu. Kääntäminen nähtiin mekaanisena korvaamisena, jossa yhden kielen ilmaisu oli suoraviivaisesti toisinnettavissa toisen kielen ilmaisulla. Keskeinen käsite oli *ekvivalenssi* (vastaavuus), jota luokiteltiin eri tavoin (esim. Koller 1979). Tekstilingvistiikan myötä tarkastelu siirtyi laajempiin kokonaisuuksiin, myös tekstien funktioihin sekä viestintään ja viestin vastaanottajiin. Esimerkiksi Eugene Nida (1964), jonka ajattelun taustalla oli Raamatun kääntäminen, erotti muodollisen ja dynaamisen ekvivalenssin ja korosti jälkimmäisen merkitystä: muodollinen tarkka käännös on usein jäykkä, kun taas dynaamisesti ekvivalentti käännös pyrkii saamaan käännöksen vastaanottajassa saman vaikutuksen kuin lähdeteksti on saanut vastaanottajassaan, ja tällainen käännös voi vaatia suuriakin kulttuurisia muutoksia. *Ekvivalenssi* on käännöstieteen keskustelluimpia ja kiistellyimpiä käsitteitä, jota myös uudempi tutkimus on lähestynyt eri näkökulmista (ks. esim. Pym 2010).

Käännöstieteen kehityksessä omaksi tieteenalaksi tärkeä tekijä oli James Holmesin (2000 [1972]) alun perin 1972 esittämä systemaattinen näkemys tieteenalasta nimeltä *Translation Studies*, sen rakenteesta ja osa-alueista (perusjakona *descriptive translation studies* ja *theoretical translation studies* tai käännösteoria, myöhemmin näistä "puhtaan" käännöstieteen suunnista erotettiin *applied translation studies*). Merkittävää oli niin ikään painopisteen siirtyminen lähdeteksteistä käännöksiin; käännösorientoituneisuutta edustavat käännöstieteessä edelleenkin keskeiset suuntaukset: funktionaaliset käännösteoriat ja deskriptiivinen käännöstutkimus.

1980-luvulla syntyneet funktionaaliset käännösteoriat, kuten Hans Vermeerin skopos-teoria (Reiß & Vermeer 1984) ja Justa Holz-Mänttärin

translatorisen toiminnan teoria (1984), korostavat kääntämistä tavoitteellisena toimintana, jossa ensisijaista on sen tarkoitus käännöksen vastaanottaja mukaan lukien. Kääntäjä on ammattimainen toimija, joka ei vain koodaa uudelleen sanoja vaan tuottaa tekstejä kulttuurienväliseen viestintään kulloisenkin toimeksiannon ja tilanteen vaatimusten mukaisesti. Siten käännöksen laatimiseen vaikuttavat muutkin seikat kuin lähdeteksti, esimerkiksi kohdekulttuurin lait, normit ja konventiot. Lisäksi translatoriseen toimintaan liittyy paljon muutakin kuin varsinainen kääntäminen: tiedonhankinta, tekstien tarkistus ja yhteistyö asiakkaiden ja muiden tekstin tuottajien kanssa. Tekstien multimodaalisuus, joka vasta myöhemmässä tutkimuksessa on tullut tärkeäksi huomioon otettavaksi näkökulmaksi, oli mukana jo Holz-Mänttärin teoriassa. Funktionalistit tulivat kielitieteen piiristä, ja heidän tutkimuksensa kohdistuivat erityisesti erikoisalojen ja käyttötekstien kääntämiseen, vaikka esimerkiksi Vermeer halusi luoda yleisen käännösteorian, joka koskee kaikkea kääntämistä.

Deskriptiivinen käännöstutkimus, jonka keskeisiä vaikuttajia oli Gideon Toury (1995), on puolestaan lähtöisin kirjallisuudentutkimuksen puolelta. Touryn kiinnostuksensa kohteena oli alun perin kirjallisuuden kääntäminen. Esimerkiksi polysysteemiteoria on kiinnostunut käännösten asemasta kohdekulttuurin kirjallisuuden systeemien verkostossa (Even-Zohar 1990). Deskriptiivisellä käännöstutkimuksella on merkittävä sija nykyisessä käännöstieteellisessä tutkimuksessa tekstilajeihin katsomatta. Kun funktionalistit pyrkivät antamaan kriteereitä hyvälle käännökselle, deskriptiivisessä käännöstutkimuksessa tavoitteena on kuvata, millaisia käännökset ovat, ja selittää, miksi ne ovat sellaisia kuin ovat. Deskriptiivisen käännöstutkimuksen tutkimuskohteita ovat olleet muun muassa kääntämisen normit ja lainalaisuudet, ns. käännösuniversaalit (ks. tarkemmin luku 3.2.).

1980-luvulta lähtien käännöstiede on nopeasti kehittynyt ja laajentunut. 2000-luvun merkittäviä suuntauksia on muun muassa kääntämisen sosiologia, joka kiinnittää huomion kääntäjään ja toimintaympäristöön. Tutkimuskohteena on myös kääntämisen (ja tulkkauksen) historia. Kognitiivinen käännöstiede puolestaan tutkii käännösprosessia, sitä mitä kääntäjän päässä tapahtuu, kun hän kääntää. Viime aikojen sosiokognitiivinen suuntaus, joka näkee kognition kehollisena ja tilannesidonnaisena,

tarkastelee käännösprosessin tutkimuksessa ”laajennettua kääntämistä” (engl. *extended translation*) ottaen huomioon kääntäjien teknologisen ja sosiaalisen työskentely-ympäristön, myös ergonomian (Risku, Rogl & Milosevic 2017). Teknologian kehityksen myötä omiksi ja kasvaviksi alueiksi ovat muodostuneet av-kääntämisen tutkimus ja käännösteknologia. Multimodaalisuuden kysymykset ja intersemioottinen kääntäminen ovat niin ikään kasvavia alueita. Nykyisissä monikielisissä yhteiskunnissa ylipäätään kysymys siitä, mikä tekee tekstistä käännöksen, mikä on lähdeteksti ja kuka on kääntäjä, ovat yhä monimutkaisempia, mikä vaatii tutkimukselta uusia lähestymistapoja, aineistoja ja menetelmiä. Käännöstieteen alueellinen laajentuminen on nostanut esiin myös sen, että länsimainen käsitys käännöksistä ja kääntämisestä on yksipuolinen ja suppea.

## 2.2. Tulkkauksen tutkimus

Koska puhuttu kieli ja keskustelu ovat ensisijaisia vuorovaikutuksen keinoja, suullista tulkkausta on esiintynyt paljon ennen kirjallista kääntämistä. Tulkkauksen systemaattinen tutkimus on kuitenkin alkanut myöhemmin kuin kääntämisen tutkimus. Tulkkauksen tutkimuksen yhtenä merkittävimpänä pioneerina on pidetty Danica Seleskovitchia, jonka ensimmäiset konferenssitulkkausta koskevat julkaisut ilmestyivät 1960-luvulla. Hänen kehittämänsä tulkinnan teoria (ra. *théorie du sens*) näkee tulkkauksen sanoman ymmärtämisenä ja tulkitsemisena; erityisen tärkeää on tulkin yleistieto ja tieto tulkattavasta asiasta (Seleskovitch & Lederer 1989). Teoria nojautui psykologiaan ja poikkesi selvästi kääntämisen puolella vallinneesta lingvistisestä lähestymistavasta, jossa sanoja koodattiin uudelleen kielestä toiseen.

Tulkkauksen tutkimuksen kohteena oli aluksi yksinomaan konferenssitulkkaus, ja ammatilliset ja koulutuskysymykset olivat keskeisiä. Tutkijat, kuten Seleskovitch, olivat itse konferenssitulkkeja. Pitkälti intuitioon perustuva tutkimus oli preskriptiivistä; siinä annettiin ohjeita tulkkien toiminnalle. (Gile 2009.) Empiirinen ja kokeellinen tutkimus alkoi 1990-luvulla. Kohteena oli edelleen konferenssitulkkaus, mutta tutkimukseen vaikuttivat keskeisesti kognitiivinen psykologia ja

neurotieteet sekä näitä lähellä olevat kielitieteen haarat psykolingvistiikka ja neurolingvistiikka (ks. Psyko- ja neurolingvistiikka tk.). Tutkijoita kiinnostivat erityisesti kognitiiviset kysymykset, kuten informaation prosessointi, kahdella kielellä tapahtuva puheen samanaikainen vastaanottaminen ja tuottaminen, ennakointi, muisti, kognitiivinen kuormitus, tulkkausstrategiat (Gile 1997) ja aivotoiminta (Tommola, Laine, Sunnari & Rinne 2000) (ks. myös Englund Dimitrova & Hyltenstam 2000).

Käännöstutkimuksen puolella jalansijaa saaneet tekstintutkimus, funktionaaliset teoriat ja kohdetekstiorientoituneisuus vaikuttivat myös tulkkauksen tutkimuksessa. Pöchhacker (2004) on kutsunut näitä suuntauksia, joille yhteistä on kiinnostus tekstuaaliseen tuotteeseen, TT-paradigmaksi (engl. *target-text-oriented translation-theoretical*). Konferenssitulkkauksesta tutkimus laajeni vähitellen dialogitulkkaukseen.

Vasta 1990-luvun lopussa ja 2000-luvulla tutkimus kiinnostui siitä tulkkauksesta, joka tapahtuu kansallisessa kontekstissa erilaisissa sosiaalisissa tilanteissa, kuten asioimistulkkauksesta eli tulkkauksesta viranomaisasioinnissa ja terveydenhuollossa sekä oikeustulkkauksesta. Oikeustulkkauksen tutkimuksessa keskeisiä kysymyksiä ovat olleet muun muassa tarkkuus ja tulkin puolueettomuus. Kun sekä oikeus- että asioimistulkkauksessa kyse on vuorovaikutuksesta, tulkkivälitteinen vuorovaikutus ja siihen liittyvät kysymykset, kuten tulkin rooli, ovat nousseet tutkimuksen keskiöön (mm. Wadensjö 1998, ks. myös luku 3.3.). (Tulkkauksen tutkimuksen historiasta ja painopisteistä ks. Pöchhacker 2004; Rouhe 2017.)

Vaikka kääntämisen ja tulkkauksen tutkimuksen perinteet ja lähtökohdat eroavat, 1990-luvulta lähtien on myös etsitty yhteistä pohjaa. Saksalaisella kielialueella vaikutusvaltaiset funktionaaliset käännösteoriat saivat tulkkauksen tutkimuksessa huomion kiinnittymään paitsi kokonaistilanteeseen ja toimintaan myös tekstilingvistisiin kysymyksiin. Käännöstieteessä vaikuttava sosiologinen suunta näkyy myös tulkkauksen tutkimuksessa; kohteena voi olla esimerkiksi tulkin status tai tulkkauksen historia. Deskriptiivisessä käännöstieteessä ja käännettyjen tekstien tutkimuksessa suosittu korpustutkimus on saanut jalansijaa myös tulkkauksen tutkimuksessa. Yhteisiä kysymyksiä ovat myös koulutukseen liittyvät ja ammatilliset kysymykset sekä laadun määrittely.

# 3. Käännöstieteen tutkimuskohteita, metodeja ja aineistoja

## 3.1. Tekstintutkimus: käännökset ja tulkkeet

Monen maan kirjallisuuden historia alkaa käännöskirjallisuudesta, ja käännöstieteellisen tutkimuksen historiallinen ydin on tekstintutkimusta. Käännettyjen tekstien osalta nykytutkimus koskee tekstilajien koko kirjoa uskonnollisten tekstien käännöksistä käännettyihin verkkoteksteihin, kaunokirjallisuudesta käyttöohjeisiin, erikoisalojen teksteistä lastenkirjallisuuteen, laki- ja säädösteksteistä tv-tekstityksiin, tuotepakkausselostuksista dubbaukseen, tietokonepelien käännöksistä oopperatekstityksiin, laulujen sanoituksista sarjakuviin, käännetystä kielimaisemasta lääkepakkausteksteihin jne. Ei liene tekstilajia, jota ei olisi käännetty. Tutkijat ovat perinteisesti koostaneet omiin tutkimustarkoituksiinsa tarkastelemansa tekstilajin käännöksistä aineistoja ja analysoineet niitä manuaalisesti, ja nämä aineistonkeruu- ja analyysimenetelmät ovat edelleen ajankohtaisia ja taajaan käytössä. Oman tutkimussuuntauksensa muodostaa korpuslingvistinen käännettyjen tekstien tutkimus, jossa tutkittava tekstiaineisto on saatettu sähköiseen muotoon. Sähköisten korpusten tutkimusta esitellään erikseen seuraavassa luvussa 3.2. Sähköisiä ja ei-sähköisiä aineistoja käytetään usein myös rinnakkain.

Käännettyjä tekstejä ei tyypillisesti tutkita sellaisenaan, vaan tutkija valitsee jonkin yksittäispiirteen tai ilmiön ja tarkastelee sen esiintymistä käännetyissä teksteissä. Tekstilajista riippuen tutkimus voi kohdistua eri ilmiöihin; esimerkkeinä mainittakoon mm. erisnimien, kulttuurispesifisten ilmausten eli reaalioiden, varieteettien, monikielisyyden, sanaleikkien, tabusanojen, huumorin, tyylipiirteiden, kieliopin ilmiöiden tai lähettäjä–vastaanottaja-konstellaatioiden kääntäminen. Käännöstieteellinen tutkimus tuottaa paljon uutta tietoa näistä ilmiöistä, mutta varsinainen tavoite on sen kuvaaminen, mitä ikkunoita käännökset avaavat kääntämiseen, kääntäjän työhön tai kohdekulttuuriin.

Menetelmällisesti ei-sähköisten tekstiaineistojen hyödyntäminen tarkoittaa ilmiöiden manuaalista poimintaa teksteistä, luokittelua ja

mahdollisesti myös muiden erilaisten analyysimenetelmien käyttöä (esim. lähiluku, diskurssianalyysi, keskustelunanalyysi, kategoria-analyysi, sisällönanalyysi, retoriikan ja tekstilingvistiikan teoriat, kieliopin erilaiset kuvausmallit jne.). Keskeinen menetelmä on vertailu: kääntäjän käännösratkaisuja vertaillaan tyypillisesti lähdetekstiin, toisenkielisiin käännöksiin tai aikaisemmin tai myöhemmin ilmestyneisiin samankielisiin käännöksiin. Julkaistut käännökset eivät heijastele vain kääntäjän ratkaisuja vaan myös kustannustoimittajan, kielentarkastajan tai muiden mahdollisten toimijoiden ”sormenjälkiä”. Mahdollista onkin vertailla julkaistua käännöstä myös julkaisemattomiin käännösversioihin ja tarkastajien kommentteihin. Vertailun menetelmiä kielentutkimuksen eri alueilla käsitellään kattavasti Kolehmaisen, Miestamon ja Nordlundin (2013) toimittamassa teoksessa.

Vertailu lähdetekstiin mahdollistaa yhtäläisyyksien ja erojen paikantamisen, ja sen avulla on mahdollista kuvata käännösstrategioita, kääntäjän kohdekielisiä ratkaisuja lähdetekstin ilmiöihin. Käännösstrategiat heijastavat kääntämisessä ja kohdekulttuurissa kulloinkin vallitsevia arvoja, ideologioita ja normeja. Tutkimuksessa on esimerkiksi kysytty, miksi puhekielen piirteet, tietyt kerronnan keinot, kirosanat, uskonnolliset tai poliittiset viittaukset ja seksuaalisten vähemmistöjen esittäminen saattavat kadota tai muuttua käännöksissä. (Suomalaisesta kääntämisen normien ja ideologioiden tutkimuksesta ks. esim. Koskinen & Paloposki 2015; Kujamäki 1998; Kuusi 2011; Tiittula & Nuolijärvi 2013.) Käännösversioiden keskinäinen vertailu taas avaa näkökulmia kääntäjän työskentelyyn ja käännösprosessiin, ja opiskelijakäännösten vertailu ammattilaisten tuotoksiin tuo tietoa kääntäjäkompetenssin kehittymisestä sekä sen luonteesta (ks. luku 3.3.). Multimodaalisten lähdetekstien vertailu niiden käännöksiin tuo tietoa kääntäjän työstä, siitä, miten kielen ulkopuoliset seikat, kuten kuvan tai äänen ja sanan suhde, vaikuttavat käännettyyn lopputulokseen esimerkiksi lasten kuvakirjojen, verkkotekstien tai tv-sarjojen ja elokuvien käännöksissä. Käännettyjä tekstejä on vertailtu lähdeteksteihin myös laadun näkökulmasta. Ajankohtainen tutkimus korostaa kuitenkin, että laatu ei ole ainoastaan käännöksen vaan myös käännösprosessin ominaisuus. Se, millaisin reunaehdoin, millaisissa ympäristöissä ja tuotantoketjuissa käännöksiä tuotetaan, vaikuttaa käännöksen laatuun (ks. esim. Abdallah 2012).

Käännöstieteellinen tekstintutkimus voi kohdistua myös teksteihin, jotka eivät ole käännöksiä. Tällaisia tekstejä ovat esimerkiksi käännöksiä ja kääntämistä kommentoivat paratekstit: ne ovat joko teoksen yhteydessä esiintyviä teosta kommentoivia tekstejä (esim. otsikko, esipuhe, jälkisanat, viitteet) tai teosta siitä erillään kommentoivia tekstejä (esim. päiväkirjamerkinnät, kirja-arviot, kirjeenvaihto, haastattelut). Niiden avulla on tutkittu esimerkiksi kääntämisen ideologiaa, käännösprosessia sekä kääntäjien verkostoja ja toimintaa (ks. esim. Kemppanen 2011; Paloposki 2007).

Tulkkausaineistojen luonne poikkeaa osin käännetyistä aineistoista. Suullinen tulkkaus on tyypillisesti ainutkertainen tilanne, jossa syntyvään tulkkeeseen tulkin ei ole mahdollista palata: tulketta ei voi jälkikäteen hioa, korjata tai muokata. Tulkkauksen tutkimuksessa tulke toimiikin tutkijalle usein ikkunana tulkkausprosessin kognitiiviseen, psykolingvistiseen ja sosiaaliseen ulottuvuuteen: tulkin muistiin, tulkkauksen kuormittavuustekijöihin sekä tulkin toimintaan välittäjänä eri osapuolia sisältävissä tilanteissa. Tulkeaineistoja on taltioitu erilaisista aidoista tulkkaustilanteista tulkkauslajeittain (simultaani-, konsekutiivi-, konferenssi-, oikeus-, asioimistulkkaus jne.) mutta myös tutkimusta varten järjestetyistä kokeellisista tulkkaustilanteista ja tulkkiopiskelijoiden oppimistilanteista.

Tulkkauksessa analyysi on kohdistunut lähdekielisen puheen ominaisuuksiin sekä siihen, kuinka lähdepuhe vaikuttaa tulkkeeseen (kieltenväliset erot, puheen ja puhujien piirteet). Tutkijat ovat tarkastelleet esimerkiksi tulkkeessa esiintyviä toistoja, korjauksia, lisäyksiä, poistoja, viiveitä ja taukoja sekä niiden syitä. Tutkittujen piirteiden avulla on lähestytty esimerkiksi tulkkausstrategioita, tulkkauksen laatua, tarkkuutta ja sujuvuutta, tulkkausprosessia, tulkin roolia ja teknisten ratkaisujen vaikutusta (esim. etätulkkauksessa). (Suomalaisesta tutkimuksesta ks. esim. Isolahti 2014; Rouhe 2017; Vik-Tuovinen 2006; Vuorikoski 2004; tulkkauksen tutkimuksesta ks. myös luku 3.3.)

Sekä käännökset että tulkkeet ovat vain yksi osa käännös- tai tulkkausprosessia. Molempiin prosesseihin vaikuttavat esimerkiksi toimeksianto, vuorovaikutustilanne, valmistautuminen ja toimijoiden etukäteen saama valmistautumismateriaali. Koko kääntämis- tai tulkkausprosessi ei siten ole tutkijan saavutettavissa pelkän käännöksen tai tulkkeen analyysin avulla.

**Käännettyjen tekstien ja tekstilajien tutkimuksesta:**

Kinnunen, Tuija. 2012. Käännöstiede. Julkaisussa: Heikkinen, Vesa, Voutilainen, Eero, Lauerma, Petri, Tiililä, Ulla & Lounela, Mikko (toim.) *Genreanalyysi – tekstilajitutkimuksen käsikirja*. Helsinki: Gaudeamus Helsinki University Press, 581–586.

Saldanha, Gabriela & O'Brien, Sharon. 2013. *Research Methodologies in Translation Studies*. London: Routledge.

**Tulkkeiden tutkimuksesta ja tulkkauksen historiasta:**

Pöchhacker, Franz. 2004. *Introducing Interpreting Studies*. London: Routledge.

Rouhe, Olga. 2017. *Kieliopin rakenne-erojen vaikutus suomi–venäjä-simultaanitulkkauksen sujuvuuteen*. Publications of the University of Eastern Finland. Dissertations in Education, Humanities, and Theology 102. Joensuu: Itä-Suomen yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-952-61-2434-6.

Vuorikoski, Anna-Riitta. 2004. *A Voice of its Citizens or a Modern Tower of Babel? The Quality of Interpreting as a Function of Political Rhetoric in the European Parliament*. Acta Universitatis Tamperensis 985. Tampere: Tampereen yliopisto. Saatavissa: http://tampub.uta.fi/bitstream/handle/10024/67348/951-44-5878-8.pdf?sequence=1&isAllowed=y.

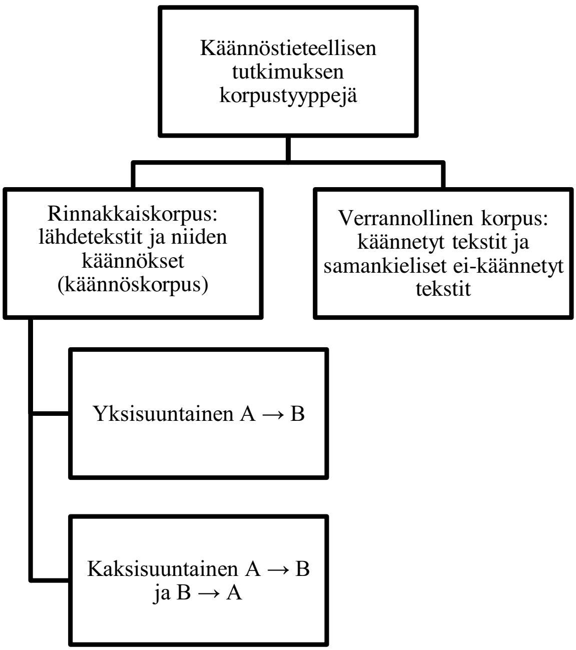
**Vertailusta tutkimusmenetelmänä:**

Kolehmainen, Leena. 2013. Käännöstiede ja kontaktilingvistiikka kohtaavat. Julkaisussa: Kolehmainen, Leena, Miestamo, Matti & Nordlund, Taru (toim.) *Kielten vertailun metodiikka*. Suomalaisen Kirjallisuuden Seuran Toimituksia 1387. Helsinki: Suomalaisen Kirjallisuuden Seura, 420–460.

Kujamäki, Pekka. 2013. Vertaileva kääntäjä, vertaileva tutkija: kieltenvälinen vertailu käännöstieteessä. Julkaisussa: Kolehmainen, Leena, Miestamo, Matti & Nordlund, Taru (toim.) *Kielten vertailun metodiikka*. Suomalaisen Kirjallisuuden Seuran Toimituksia 1387. Helsinki: Suomalaisen Kirjallisuuden Seura, 355–395.

## 3.2. Korpuslingvistinen käännösten ja tulkkeiden tutkimus

Edellisessä luvussa havainnollistetut käännöksiä tai tulkkeita sisältävät aineistot on mahdollista saattaa sähköiseen muotoon, jolloin niitä voi analysoida erilaisten automaattisten työkalujen ja ohjelmien avulla. Suuria sähköisiä kirjoitetun tai puhutun kielen aineistoja kutsutaan yleensä *korpuksiksi* (ks. Korpusaineistot tk.). Kuva 1 havainnollistaa käännöstieteellisen tutkimuksen tärkeimpiä korpustyyppejä. Keskeiset termit, *rinnakkaiskorpus* (engl. *parallel corpus*) ja sen synonyyminä esiintyvä *käännöskorpus* (engl. *translation corpus*) sekä *verrannollinen korpus* (engl. *comparable corpus*), ovat pitkälti vakiintuneet, mutta on syytä olla tietoinen siitä, että termien käyttö ja merkitys vaihtelevat eri tutkijoiden kesken. Havainnollistamme seuraavassa näitä aineistoja ja niiden avulla tutkittuja kysymyksiä.




Kuva 1. Käännöstieteellisen tutkimuksen keskeiset korpustyypit.

*Rinnakkaiskorpuksilla* eli *käännöskorpuksilla* tarkoitetaan yleensä kaksi- tai monikielisiä korpuksia, jotka koostuvat yhden kielen lähdeteksteistä ja niiden käännöksistä yhteen tai useampaan kieleen. Mahdollisesti mukana voi olla saman lähdetekstin useampia käännöksiä samaan tai eri kieliin. Rinnakkaiskorpus voi olla yksisuuntainen ja sisältää käännöksiä vain yhteen suuntaan tai kaksisuuntainen, jolloin mukana on käännöksiä kielestä A kieleen B ja kielestä B kieleen A. Kaksisuuntaisten rinnakkaiskorpusten pioneerina pidetään usein englanti–norja-rinnakkaiskorpusta[1] (Johansson 2007). Se koostuu englanninkielisistä lähdeteksteistä ja niiden norjannoksista sekä norjankielisistä lähdeteksteistä ja niiden englanninnoksista. Korpus mahdollistaa siten kieltenvälisen vertailun keskenään rinnakkaisten tekstien välillä (lähdetekstien vertailu käännöksiin) molemmissa suunnissa (englanti–norja, norja–englanti) ja keskenään verrannollisten tekstien välillä (englanninkieliset vs. norjankieliset alkuperäistekstit, englanninkieliset vs. norjankieliset käännökset). Lisämahdollisuutena on yksikielinen alkuperäistekstien vertailu samankielisiin käännöksiin (englanninkieliset alkuperäistekstit vs. käännökset, norjankieliset alkuperäistekstit vs. käännökset).

Verrannolliset korpukset puolestaan koostuvat teksteistä, jotka eivät ole toistensa käännöksiä mutta ovat jollakin muulla tapaa vertailukelpoisia keskenään, esimerkiksi tekstilajiltaan. Käännöstieteessä verrannollisia korpuksia voidaan pitää 1990-luvulla alkaneen korpuslingvistisen "buumin" yhtenä laukaisijana. Idea verrannollisista korpuksista palautuu Bakerin (1993; 1996) artikkeleihin, joissa esiteltiin kokonaan uudenlainen aineistoasetelma: edellä esitellyn rinnakkaiskorpusten mahdollistaman lähdetekstien ja niiden käännösten vertailun sijaan käännettyjä tekstejä olisikin mahdollista vertailla myös samankielisiin ei-käännettyihin teksteihin. Verrannollisten korpusten tutkimuksessa alettiin keskittyä enemmän kohdeteksteihin, kun aiemmin keskityttiin lähdeteksteihin. Tunnettuja verrannollisia korpuksia ovat esimerkiksi Translational English Corpus[2] (Laviosa 1996), joka sisältää eri kielistä englantiin käännettyjä ja verrannollisia ei-käännettyjä englanninkielisiä tekstejä, ja Käännössuomen korpus (Mauranen 2000), joka koostuu eri kielistä suomeen käännetyistä ja ei-käännetyistä, ns. supisuomen teksteistä. Suomalainen YLE-korpus, joka lienee edelleenkin

maailman suurin ruututekstikorpus, lukeutuu verrannollisiin korpuksiin. Se sisältää Ylen vuosina 1992–2004 lähettämien tv-ohjelmien eri kielistä käännetyt tekstitykset ja kuulovammaisille katsojille suunnatut ei-käännetyt ohjelmatekstitykset.[3] Tiedemannin Opus on rinnakkaiskorpus, joka käsittää internetistä kerättyjä käännöstekstejä, myös elokuvien ja tv-ohjelmien tekstityksiä.[4]

Korpuslingvistiset tutkimukset ovat pitkälti keskittyneet mahdollisten kääntämisen universaalien selvittämiseen. Hypoteesi kääntämisen universaaleista, kielellisistä piirteistä, jotka esiintyvät kieliparista riippumatta pikemminkin käännetyissä teksteissä kuin ei-käännetyissä ja joita kääntäminen erityisenä kielentuottamisprosessinaan aiheuttaa kohdeteksteihin, palautuu Bakerin (1993; 1996) julkaisuihin. Kääntämisen universaaleja tai lakeja, jotka erottavat käännöksiä samankielisistä ei-käännetyistä teksteistä (Toury 1995), on tutkittu korpuksissa tarkastelemalla esimerkiksi sana- ja sanemääriä ja leksikaalista variaatiota, mittaamalla lause- ja virkepituuksia, tutkimalla kollokaatioita (eli mitkä sanat esiintyvät tyypillisesti yhdessä) tai analysoimalla lukuisten erilaisten kielen yksiköiden, kuten pronominien, konjunktioiden, fraseologismien tai kieliopillisten rakenteiden, käyttöympäristöjä, funktioita ja esiintymistiheyksiä.

Kääntämisen universaaleiksi on ehdotettu mm. käännösten kielellistä eksplisiittistymistä (tendenssi lisätä käännetyn tekstin eksplisiittisyyttä esim. koheesiokytkennöillä ja selityksillä), yksinkertaistumista ja normaalistumista (viimeisintä kutsutaan vaihtoehtoisesti myös konventionaalistumiseksi, standardisoitumiseksi tai tasaantumiseksi). Kollokaatioiden on havaittu olevan epätyypillisiä käännöksissä, ja kielellisten yksiköiden frekvenssien on huomattu olevan poikkeuksellisia. Esimerkiksi kieliparikohtaiset kohdekielen ns. vastineettomat ainekset näyttävät olevan harvinaisempia kuin ei-käännetyissä teksteissä tai puuttuvan kokonaan.[5]

Kääntämisen universaalien tutkimuksessa rinnakkaiskorpusten käyttö on jatkunut verrannollisten korpusten rinnalla. Tämä on aiheuttanut joskus käsitteellistä sekaannusta, ja aina ei ole täysin selvää, mitä tutkijat täsmälleen tarkoittavat. Chestermanin (2004) erottelu S- ja T-universaalien välille onkin selkiyttänyt keskustelua: S-universaalit (S = *source*)

viittaavat käännettyjen tekstien piirteisiin verrattuna niiden lähdeteksteihin. T-universaalit (T = *target*) taas tarkoittavat systemaattisia eroja kohdekielisten käännösten ja samankielisten ei-käännettyjen tekstien välillä. Kääntämisen universaalihypoteeseja on myös kritisoitu, ja kaikille ehdotetuille hypoteeseille on löydetty vastaesimerkkejä korpuksista, muita selityksiä tai vasta-argumentteja kääntämisen historiasta (ks. esim. Becher 2011; Kuusi 2016; Tiittula & Nuolijärvi 2016).

Kääntämisen universaalien tutkimuksessa tavoitteena oli alun perin selvittää, mikä on käännetyille teksteille yhteistä lähdekielestä riippumatta ja mitä piirteitä kääntäminen omana erityisenä kielen tuottamisprosessinaan aiheuttaa. Mukana tarkastelussa on ollut kuitenkin alusta asti myös kysymys kieliparikohtaisista tai kieltenvälisistä typologisista eroista ja niiden vaikutuksesta käännettyjen tekstien kielellisiin piirteisiin (ks. esim. Cappelle & Loock 2017), interferenssistä eli lähdekieli- tai lähdetekstispesifisestä kontaktivaikutuksesta käännettyihin teksteihin (ks. esim. Mauranen 2004) sekä käännössuunnan vaikutuksesta (ks. esim. Evert & Neumann 2017). Tutkijat ovat myös kysyneet, mikä on tekstilajikohtaista, mikä kääntämiselle universaalia tekstilajista riippumatta (ks. esim. Delaere & De Sutter 2017). Lisäksi korpustutkijat ovat nostaneet esille muiden mahdollisten toimijoiden käännöksiin jättämät jäljet – kääntäminen on sosiaalinen tapahtuma, jossa lopputulokseen vaikuttavat kääntäjien lisäksi kustannustoimittajat, editoijat ja kielentarkastajat. Korpuksiin taltioidut käännetyt tekstit eivät siten välttämättä kerro vain kääntäjän ratkaisuista. (Kruger 2017.) Vaihtelua aiheuttaa myös se, kenen käännöksestä on kyse. Kääntäminen voi tapahtua kääntäjän äidinkieleen tai vieraaseen kieleen – ja aina ei ole edes selvää, onko kyse äidinkielestä vai vieraasta kielestä (ks. esim. Pokorn 2005). Yhteenvetona kääntämisen universaalipiirteiden tutkimuksesta voi todeta, että nykykorpustutkimuksen valossa käännettyjen tekstien mahdolliset erityisyydet näyttäytyvät monimutkaisina piirteiden kimppuina, jotka voivat johtua useasta samanaikaisesti vaikuttavasta tekijästä, ei pelkästään käännösprosessista.

Korpusten avulla on tutkittu myös tekstien ideologiaa: keskeisenä menetelmänä tässä tutkimuksessa on ollut avainsana-analyysi. *Avainsanalla* tarkoitetaan kielenainesta, jonka esiintymistiheys aineistossa poikkeaa

tilastollisesti merkitsevästi vertailuaineistosta. Poikkeuksellinen esiintymistiheys kertoo esimerkiksi käännettyjen tekstien näkökulmapiirteistä ja diskursiivisista erityispiirteistä tai eroista vertailuaineistoihin verrattuna (ks. esim. Kemppanen 2008; Probirskaja 2009; Laviosa, Pagano, Kemppanen & Ji 2017). Automaattisten korpuslingvististen menetelmien rinnalla tutkijat hyödyntävät usein tilastollisia menetelmiä (ks. Määrällinen korpuslingvistiikka tk.).

Kielenoppimisen tutkimuksessa koostettuihin oppijankielen korpuksiin taas rinnastuvat käännöstieteilijöiden opiskelijakäännöskorpukset, joita esitellään esimerkiksi Castagnolin, Ciobanun, Kunzin, Küblerin ja Volanschin (2011) artikkelissa. Alan terminologia ei ole vielä vakiintunut vaan vaihtelee (engl. *translation learner corpus, learner translator corpus, student translation corpus*). Opiskelijakäännöskorpusten avulla on mahdollista tutkia esimerkiksi opiskelijoiden käännöskompetenssia ja sen kehittymistä ja vertailla sitä ammattikääntäjien kompetenssiin (ks. esim. Englund Dimitrova 2005).

Sähköiset tulkkauskorpukset on mahdollista jakaa samalla tavalla lähdetekstejä ja niiden tulkkeita sisältäviin kaksi- tai monikielisiin rinnakkaiskorpuksiin sekä verrannollisiin tulkkausta ja samankielistä ei-tulkattua vuorovaikutusta sisältäviin korpuksiin (ks. esim. Bendazzoli 2015; Bendazzoli 2018; Bernardini, Ferraresi, Russo, Collard & Defrancq 2018; Shlesinger 1998). Shlesinger (1998, 3) korostaa, että ideaalitilanteessa molemmat tulkkauskorpukset sisältävät lisäksi kohdekielen käännettyjä tekstejä. Tällainen korpuskonstellaatio mahdollistaa sen tunnistamisen, mikä on nimenomaan tulkkaukselle tyypillistä ja erottaa sitä kääntämisestä. Tulkkauskorpukset eivät ole kehittyneet samaan tahtiin kirjoitetun kielen käännöksiä sisältävien korpusten kanssa, niitä ei ole saatavilla yhtä paljon, eivätkä ne ole yhtä suuria. Syy on ilmeinen: suullisen vuorovaikutuksen video- tai äänitaltioiminen, puheen litterointi ja paralingvististen keinojen merkintä litteraattiin on työläämpää kuin kirjoitettujen tekstien työstäminen.

Tulkkausmenetelmä ja -tilanne ovat keskeisimmät seikat, jotka vaikuttavat tulkkaukseen. Tästä syystä tulkkauskorpukset edustavat usein kulloinkin vain yhtä tai muutamaa tulkkausmenetelmää ja -tilannetta. Esimerkkeinä olemassa olevista sähköisistä tulkkauskorpuksista

mainittakoon Euroopan parlamentin kokousten englanti–italia–espanja-simultaanitulkkauskorpus EPIC (The European Parliament Interpreting Corpus)[6] (Monti, Bendazzoli, Sandrelli & Russo 2005), monikielinen ComInDat (Community Interpreting Database)[7], joka sisältää sekä lääkäri–potilas-asioimistulkkausta että oikeustulkkausta (Angermeyer, Meyer & Schmidt 2012), ja englanti–italia-simultaanitulkkauskorpus FOOTIE (Football in Europe) jalkapallon lehdistökonferensseista (Sandrelli 2012).

Käännösten korpuslingvistiseen tutkimukseen on saatavilla lukuisia johdanto- ja kokoomateoksia.

**Ajankohtaisia johdantoteoksia käännöstieteellisten korpusten kokoamisesta, annotoinnista ja korpuslingvistisestä analyysista:**

Mikhailov, Mikhail & Cooper, Robert. 2016. *Corpus Linguistics for Translation and Contrastive Studies: A Guide for Research*. Abingdon: Taylor & Francis.

Zanettin, Federico. 2012. *Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*. Manchester: St. Jerome.

**Ajankohtaisia kokoomateoksia käännösten korpuslingvistisestä tutkimuksesta:**

Delaere, Isabelle, Lefer, Marie-Aude & De Sutter, Gert (toim.). 2017. *Empirical Translation Studies: New Methodological and Theoretical Traditions*. Berlin: De Gruyter Mouton.

Hansen-Schirra, Silvia, Neumann, Stella & Steiner, Erich (toim.). 2012. *Cross-Linguistic Corpora for the Study of Translations. Insights from the Language Pair English-German*. Berlin: De Gruyter Mouton.

Ji, Meng (toim.). 2016. *Empirical Translation Studies: Interdisciplinary Methodologies Explored*. Sheffield: Equinox.

Ji, Meng, Hareide, Lidun, Li, Defeng & Oakes, Michael (toim.). 2017. *Corpus Methodologies Explained: An Empirical Approach to Translation Studies*. London: Routledge.

Kruger, Alet, Wallmach, Kim & Munday, Jeremy (toim.). 2013. *Corpus-based Translation Studies: Research and Applications*. London: Continuum.

Laviosa, Sara, Pagano, Adriana, Kemppanen, Hannu & Ji, Meng. 2017. *Textual and Contextual Analysis in Empirical Translation Studies*. Singapore: Springer.

Oakes, Michael P. & Ji, Meng (toim.). 2012. *Quantitative Methods in Corpus-based Translation Studies: A Practical Guide to Descriptive Translation Research*. Amsterdam: John Benjamins.

**Katsauksia korpuslingvistiseen tulkkaustutkimukseen ja ajankohtaista tutkimusta:**

Bendazzoli, Claudio. 2015. Corpus-based research. Julkaisussa: Pöchhacker, Franz (toim.) *Routledge Encyclopedia of Interpreting Studies*. Florence: Taylor & Francis, 132–135.

Russo, Mariachiara, Bendazzoli, Claudio & Defrancq, Bart (toim.). 2018. *Making Way in Corpus-based Interpreting Studies*. Singapore: Springer.

Setton, Robin. 2011. Corpus-based interpreting studies (CIS). Julkaisussa: Kruger, Alet, Wallmach, Kim & Munday, Jeremy (toim.) *Corpus-based Translation Studies: Research and Applications*. London: Continuum, 33–75.

Shlesinger, Miriam. 1998. Corpus-based interpreting studies as an offshoot of corpus-based translation studies. *Meta* XLIII: 4, 1–8. Saatavissa: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.119.9115&rep=rep1&type=pdf.

Straniero Sergio, Francesco & Falbo, Caterina (toim.). 2012. *Breaking Ground in Corpus-based Interpreting Studies*. Bern: Peter Lang.

## 3.3. Prosessin tutkimus

Käännöstutkimuksessa painopiste on pitkälti ollut lopputuotteessa, käännöksissä, vaikka prosessi vaikuttaa olennaisesti käännöksen laatuun. Suuri osa tulkkauksen tutkimusta taas on ollut nimenomaan prosessin tutkimusta (vrt. luku 2.2.). Koska kirjallinen kääntäminen ja suullinen tulkkaus ovat erilaista toimintaa, myös niiden prosessien tutkimuksessa on eroja. Seuraavassa painopiste on aluksi kognitiivisessa prosessin tutkimuksessa ja erityisesti kääntämisessä, mutta otamme soveltuvissa kohdin myös tulkkauksen huomioon. Luvun lopussa on lyhyt katsaus tulkkauksen tutkimuksessa käytetyistä menetelmistä.

Prosessin tutkimuksen keskeisenä tavoitteena on pidetty kääntämisen kognitiivisten prosessien ymmärtämistä painopisteen ollessa yksittäisessä kääntäjässä. Sekä kääntämisen että tulkkauksen prosessi on tavallisesti jaettu kolmeen vaiheeseen: lähdetekstin ymmärtämiseen, kahden kielen välillä tapahtuvaan siirtoon ja kohdetekstin tuottamiseen. (Englund Dimitrova 2016.) Prosessia voidaan kuitenkin tarkastella laajemmin, ja esimerkiksi Muñoz Martín (2010) erottaa käännösprosessissa kolme tasoa: Ensimmäinen taso käsittää käännösprosessin aikaiset mentaaliset tilat ja operaatiot. Toinen, ensimmäiseen kiinteästi liittyvä taso käsittää kääntämisen osatehtävät, kuten lukemisen, kirjoittamisen, tiedonhankinnan ja oikoluvun. Kolmas taso viittaa käännösprosessin tilanteisuuteen ja sisältää kaiken käännöstoimeksiantoon ja käännökseen liittyvän toiminnan ja kaikki mukana olevat toimijat alkaen toimeksiantajan kontaktista ja päättyen lopullisen käännöksen vastaanottoon (vrt. Holz-Mänttärin 1984 translatorisen toiminnan malli). Tällä tasolla käännösprosessiin kuuluvat niin yhteistyö eri toimijoiden kanssa kuin monet ammattikääntämiseen liittyvät lisätehtävät, kuten projektinhallinta tai laskutus. Tarkastelemme tässä luvussa kahta ensimmäistä tasoa ja niitä tutkivaa kognitiivista suuntausta eli käännösprosessia suppeammassa mielessä. Kolmatta tasoa, joka laajentaa näkökulman eri toimijoihin ja toimintaympäristöön, käsittelemme seuraavassa luvussa.

Kognitiivisessa käännösprosessin tutkimuksessa huomio on kohdistunut muun muassa ongelmanratkaisuun ja päätöksentekoon sekä käännösstrategioihin. Keskeisiä kysymyksiä ovat niin ikään kompetenssi

ja asiantuntijuus sekä niiden kehittyminen, esimerkiksi miten ammattikääntäjän toiminta eroaa kokemattoman kääntäjän toiminnasta tai miten kompetenssia voidaan mallintaa. Koska kääntäminen on aina sidoksissa sosiaaliseen kontekstiin, prosessin tutkimuksessa pyritään myös ymmärtämään sen vaikutusta. (Saldanha & O'Brien 2013, 109.)

Varhaisimpia tutkimusmenetelmiä on introspektio, kääntäjän sanallinen raportti omasta prosessistaan. Raportti voi olla suullinen tai kirjallinen, ja se voidaan tuottaa tehtävää suoritettaessa tai sen jälkeen. Esimerkiksi kääntäjä voi pitää päiväkirjaa prosessin aikana tai hän voi kertoa prosessistaan käännöksen jälkeen. Jälkeenpäin tapahtuvaa retrospektiota voidaan käyttää myös tulkkauksen tutkimuksessa (esim. Vik-Tuovinen 2006; retrospektiosta käännösprosessin tutkimuksessa ks. Hansen 2006). Kognitiivisesta psykologiasta lainattu ääneen ajattelun menetelmä oli pitkään keskeinen kognitiivisen käännöstutkimuksen menetelmä. Siinä tehtävän suorittajaa pyydetään verbalisoimaan kaikki ajatuksensa tehtävän aikana ja puhe litteroidaan ”ääneenajatteluprotokollaksi” (*think-aloud protocol*, TAP). Göpferich, Bayer-Hohenwarter, Prassl ja Stadlober (2011) ovat ehdottaneet menetelmän nimeksi käännösprosessiprotokollaa (*translation process protocol*, TPP) osoittamaan sitä, että protokolla ei sisällä vain kääntämisen aikaista puhetta vaan myös toimintaa prosessin aikana, kuten tiedonhakua.

Menetelmällä on tutkittu kääntäjien ongelmanratkaisua ja eri ryhmien välisiä eroja, kuten kokeneiden kääntäjien ja opiskelijoiden käännösprosessien eroja (esim. Jääskeläinen 1999). Menetelmää on kritisoitu siitä, että ääneen ajattelu häiritsee käännöstyötä eikä ole luonnollista. Lisäksi kaikki toiminnot, erityisesti automatisoituneet ja heikosti tiedostetut, eivät ole verbalisoitavissa. Rinnalle on esitetty parityöskentelyssä syntyvää dialogista protokollaa (Kussmaul 1995). Tällöin yhdessä työskentelevät kääntäjät keskustelevat käännösongelmista ja niiden ratkaisuista, ja nauhoitettu keskustelu analysoidaan (mm. Pavlović 2013). On huomattava, että TAP ja dialoginen protokolla paitsi ovat eri menetelmiä myös kuvaavat erilaisia prosesseja (Jääskeläinen 2011b, 19).

Ääneen ajattelu -metodin rinnalle on tullut muita menetelmiä työkalujen kehittyessä. Tietokoneen näppäintoimintojen tallennusohjelmalla saadaan talteen kaikki käännösprosessin aikana tapahtuvat

kirjoitustoiminnot mukaan lukien hiiren liikkeet, tauot ja poistot. Näin voidaan tarkastella esimerkiksi sitä, miten kääntäjä jaksottaa kääntämistä sekä editoi ja korjaa tuotostaan. Käännöstieteellisessä tutkimuksessa on käytetty erityisesti Translog-ohjelmaa. Esimerkiksi tekstin segmentointia tutkittaessa on havaittu, että kokeneet kääntäjät prosessoivat laajempia yksiköitä kuin kokemattomat (Jakobsen 2003; Englund Dimitrova 2005). Taukoja analysoimalla voidaan myös tutkia kääntäjän tai post-editoijan kognitiivista kuormitusta (O'Brien 2006). Myös kuvaruuduntallennusohjelma, joka tallentaa tietokoneen näytön tapahtumat, mahdollistaa kirjoitusprosessin yksityiskohtaisen tutkimisen. Lisäksi tallennetuiksi tulevat sähköiset tiedonhaut, esimerkiksi elektronisten sanakirjojen käyttö. Kumpulainen (2016) on tutkinut menetelmän avulla kääntäjäopiskelijoiden käännösprosessia, kuten käännösyksikön pituutta, korjauksia ja tiedonhakua. Hänen pitkittäistutkimuksensa kohteena on käännöskompetenssin kehittyminen.

Silmänliike- eli katseenseurantatutkimuksen avulla voidaan tutkia kääntäjän ja tulkin kognitiivista suoritusta ja kuormitusta prosessin aikana. Esimerkiksi Jakobsen ja Jensen (2008) tutkivat sitä, miten lähdetekstin lukeminen eroaa sen mukaan, onko tarkoituksena tavallinen tekstin ymmärtäminen vai valmistautuminen tekstin kääntämiseen ja käännetäänkö suullisesti (prima vista -kääntäminen) vai kirjallisesti. Koglin (2015) puolestaan vertaili kognitiivista kuormitusta metaforien kääntämisessä ja niiden konekäännösten post-editoinnissa. Silmänliiketutkimusta voidaan soveltaa myös vaikkapa kuvailutulkkauksen tutkimuksessa sen selvittämiseen, miten näkevä tulkki identifioi ja erottaa kokonaisuudesta salientin informaation, joka on olennainen auditiivisen kertomuksen rakentamisessa sokealle vastaanottajalle (Kruger 2012).

Käännösprosessin tutkimuksessa korostetaan nykyään triangulaatiota (ks. Johdannon luku 1 tk.): koska kyse on kompleksisista prosesseista, niiden ymmärtäminen vaatii eri menetelmien yhdistämistä. Monimenetelmäisyys mahdollistaa myös uudenlaisia tutkimuskysymyksiä, kun aiemmin tutkimus uusien työkalujen sovelluksessa on ollut enemmän metodilähtöistä. Esimerkiksi yhdistämällä silmänliike- ja näppäintoimintojen tutkimusta on päästy tarkastelemaan käännösprosessin keskeisten osien, lähdetekstin ymmärtämisen ja kohdetekstin tuottamisen,

suhdetta (Dragsted 2010). Erityisesti käännöskompetenssin tutkimuksessa ja mallintamisessa monimenetelmäisyys on tavallista, ja prosessin tutkimukseen yhdistetään myös lopputulosten eli lopullisten käännösten analyysi (esim. Englund Dimitrova 2005; Göpferich ja muut 2011; Kumpulainen 2016; PACTE Group 2011).

Samalla kun kognitiivisen käännösprosessin tutkimuksen kohteet ovat laajentuneet, esiin ovat nousseet yhä enemmän metodiset kysymykset, toisin sanoen kysymykset siitä, miten prosessia ylipäätään voidaan tutkia. Myös tutkimusasetelmaan on kiinnitetty tarkempaa huomiota, esimerkiksi tutkittavien valintaan ja luokitteluun (miten esim. määritellään ammattikääntäjä). Tutkimus on usein monitieteistä, ja erityisesti kognitiotieteen ja tietojenkäsittelytieteiden asiantuntemus on tarpeen. (Kognitiivisesta käännöstutkimuksesta tarkemmin ks. esim. Shreve & Angelone 2010; Risku 2012.)

Prosessin tutkimus on usein myös kokeellista. Gilen (2015) mukaan kokeelliset menetelmät sopivat hyvin käännös- ja tulkkaustutkimukseen, koska kääntämiseen ja tulkkaukseen vaikuttavat aina monet tekijät ja niitä sekä kääntämisen ja tulkkauksen ympäristöä on mahdollista kontrolloida ja muunnella ja tuotos on helposti määriteltävissä. Usein tutkimus on eksploratiivista (vähän tutkittua aluetta kartoittavaa, pilottitutkimusta) tai kvasikokeellista (kaikkia muuttujia ei kontrolloida tai manipuloida). Koeasetelmissa voidaan tutkia esimerkiksi käännössuunnan vaikutusta prosessiin. Tulkkauksen puolella suuri osa simultaanitulkkauksen tutkimuksesta on ollut kokeellista ja hypoteesien testausta, ja tutkimus on kohdistunut muun muassa muistiin, kielen prosessointiin ja kognitiivisen kuormituksen vaikutuksiin (Hale & Napier 2013).

Kokeellisen tutkimuksen ongelma on, että tilanne ei useinkaan ole aito – paitsi ehkä opiskelijoille, joita onkin usein käytetty tutkimuksissa. Esimerkiksi tutkija valikoi käännettävän tekstin, joka voi ammattikääntäjän näkökulmasta olla epäolennainen. Jos tehtävä ei ole aito, ammattikääntäjiä on vaikea motivoida paneutumaan kääntämiseen niin kuin he työskentelisivät aidossa toimeksiannossa. Käännösprosessin tutkimuksessa kiinnitetäänkin yhä enemmän huomiota ekologiseen validiteettiin eli siihen, että koetilanne vastaa tutkittavaa todellisuutta ja kokeen tulokset ovat yleistettävissä tositilanteisiin. Myös etnografisten

menetelmien avulla pyritään parantamaan tähänastisen prosessitutkimuksen puutteita (ks. Lingvistinen etnografia toimintayhteisöissä tk.). Sosiokognitiivisessa lähestymistavassa käännösprosessin tutkimuksessa otetaan huomioon kääntäjän ja hänen ympäristönsä vuorovaikutus, johon sisältyy paitsi työkalujen käyttö myös esimerkiksi kontaktit kollegoihin ja asiakkaisiin (Risku 2014; Risku ja muut 2017). Tutkija voi videoida kääntäjää tämän työskennellessä omassa työympäristössään tai havainnoida ja tehdä muistiinpanoja. Muistiinpanojen tarkkuutta lisää, jos tutkija puhuu ne auki välittömästi havainnointitilanteen jälkeen sekä nauhoittaa ja litteroi puheensa (Risku 2014). Bundgaardin, Paulsen Christensenin ja Schjoldagerin (2016) tutkimuksessa ihmisen ja koneen välisestä vuorovaikutuksesta käytettiin makrotason analyysissa etnografisia menetelmiä (havainnointia, haastatteluja, dokumenttien analyysia) ja mikrotason analyysissa kokeellisia menetelmiä (näppäin- ja kuvaruututallennusta, haastatteluja, kyselyjä).

Käännösprosessia voidaan tutkia myös tarkastelemalla eri käännösversioita ja muita käännöksiin liittyviä dokumentteja, kuten kääntäjien päiväkirjoja, kustantajien ja kääntäjien kirjeenvaihtoa ja muuta arkistomateriaalia (Munday 2015). Tällainen tutkimus on osoittanut, että käännöksen takana voi olla hyvin monia toimijoita (Walker 2017). Kääntäjien yhteistyö ja yhteiskääntäminen ei ilmiönä ole uusi, mutta se on saanut uusia muotoja sähköisten verkkotyökalujen myötä. Esimerkkinä tästä on käännösten joukkoistaminen, jossa vapaaehtoinen verkkoyhteisö tuottaa käännöksen. Yhteiskääntämistä on tutkittu muun muassa kyselyjen ja haastattelujen avulla (ks. tarkemmin Jiménez-Crespo 2016). Mikäli käännöstä tuotetaan yhdessä samassa tilassa, prosessia voidaan tutkia myös multimodaalisen vuorovaikutuksen tutkimuksen keinoin (Hirvonen & Tiittula 2018).

Tulkkauksen tutkimuksessa Pöchhacker (2004, 64) erottaa kolme pääasiallista lähestymistapaa: kenttätutkimuksen, kyselyn ja kokeellisen tutkimuksen. Kenttätutkimuksessa aineistoa kerätään aidosta ympäristöstä havainnoimalla tai tallentamalla. Tulkkaustilanteiden tallentamiseen voi olla vaikea saada nauhoitus- ja erityisesti videointilupia, minkä vuoksi tutkija voi joutua turvautumaan pelkkiin muistiinpanoihin. Tällöin havaintojen ja muistiinpanojen systematiikkaan ja analyysiin on kiinnitettävä erityistä huomiota. Tallenteiden avulla ja diskurssi- ja

keskustelunanalyyttisin metodein voidaan analysoida tulkkausvuorovaikutusta, esimerkiksi sitä, miten tulkki koordinoi vuorovaikutusta (Wadensjö 1998) tai miten erilaiset tulkkausstrategiat vaikuttavat tulkattavien osallistumismahdollisuuksiin (Gavioli & Baraldi 2011), tai tulkin ei-verbaalin viestinnän merkitystä (Mason 2012). Lisäksi on mahdollista analysoida vuorovaikutuksen muiden osapuolten, kuten oikeustulkkauksessa tuomareiden ja asianajajien, toimintaa eli sitä, miten he työskentelevät tulkin kanssa ja tulkin välityksellä asiakkaan kanssa (Kinnunen 2010; 2013). Tavallisesta monikielisestä arkivuorovaikutuksesta ja työelämästä taltioidut tilanteet puolestaan tuovat ilmi, kuinka tulkkaus voi limittyä monikielisten osallistujien muuhun kielelliseen vuorovaikutukseen ja muihin rooleihin (Harjunpää 2017; Koskela, Koskinen & Pilke 2017). Maallikkotulkkien tulkkeiden analyysi taas nostaa esiin sen, millaisia eettisiä ja laatuongelmia kouluttamattoman tulkin käyttö voi aiheuttaa (Pöchhacker & Kadric 1999).

Aitojen ja vertailukelpoisten tulkkausaineistojen puuttuessa on käytetty simuloituja tulkkaustilanteita. Tällöin tärkeää on, että osallistujat toimivat niissä rooleissa, joissa he toimivat ammatillisesti. Esimerkiksi näin on vertailtu poliisikuulustelujen kasvokkaista tulkkausta etätulkkaukseen (Braun & Taylor 2011). Usein prosessia tutkitaan myös analysoimalla lähde- ja kohdekielisiä korpuksia (ks. luku 3.2.).

**Katsauksia käännösprosessin tutkimukseen ja metodikirjallisuutta:**

Englund Dimitrova, Birgitta. 2016. Translation process. Julkaisussa: Gambier, Yves & Doorslaer, Luc van (toim.). 2010–2016. *Handbook of Translation Studies Online*. [verkkoaineisto]. Amsterdam: John Benjamins.

Jääskeläinen, Riitta. 2011a. Studying the translation process. Julkaisussa: Malmkjær, Kirsten & Windle, Kevin (toim.) *The Oxford Handbook of Translation Studies*. Oxford: Oxford University Press, 123–135.

Risku, Hanna. 2012. Cognitive approaches to translation. Julkaisussa: Chapelle, Carol A. (toim.) *The Encyclopedia of Applied Linguistics*. [verkkoaineisto]. DOI: 10.1002/9781405198431.wbeal0145.

**Ajankohtaisia kokoomateoksia käännösprosessin tutkimuksesta:**

Alvstad, Cecilia, Englund Dimitrova, Birgitta, Hild, Adelina & Tiselius, Elisabet (toim.). 2011. *Methods and Strategies of Process Research: Integrative Approaches in Translation Studies*. Amsterdam: John Benjamins.

Ehrensberger-Dow, Maureen, Göpferich, Susanne & O'Brien, Sharon (toim.). 2015. *Interdisciplinarity in Translation and Interpreting Process Research*. Amsterdam: John Benjamins.

Muñoz Martín, Ricardo (toim.). 2014. *Minding Translation. MonTI, Special Issue 1*. Saatavissa: http://rua.ua.es/dspace/handle/10045/43719.

Muñoz Martín, Ricardo (toim.). 2016. *Reembedding Translation Process Research*. Amsterdam: John Benjamins.

O'Brien, Sharon (toim.). 2011. *Cognitive Explorations of Translation*. London: Continuum.

Shreve, Gregory M. & Angelone, Erik (toim.). 2010. *Translation and Cognition*. Amsterdam: John Benjamins.

## 3.4. Toimijoiden, toimintaympäristön ja vastaanoton tutkimus

Toimijoiden ja toimintaympäristön tutkimuksessa tavanomainen lähestymistapa on etnografinen, sillä etnografia pyrkii kuvaamaan yksityiskohtaisesti kulttuuria, yhteisöä tai ryhmää, sen instituutioita, uskomuksia, tuotteita ja käyttäytymistä sosiaalisessa kontekstissaan (ks. esim. Fetterman 2011). Ryhmän toimintaa ei tarkastella vain tutkijan näkökulmasta, vaan olennaista on yhteisöön kuuluvien näkökulma. Etnografinen tutkimusote on tullut yhä suositummaksi käännöstieteessä 2000-luvulla, kun tutkijat ovat kiinnostuneet kääntämisestä ja tulkkaamisesta sosiaalisena ja tilanteisena toimintana, kääntämisen ja tulkkauksen käytänteistä ja prosesseista, toimintaympäristöistä, toimijoista sekä kääntäjän tai tulkin roolista käännös- ja tulkkausviestinnässä.

Etnografian ydintä ovat kenttätutkimus ja yhteisön toiminnan havainnointi sen luonnollisessa ympäristössä, mutta aineisto ja sen keruun ja analyysin menetelmät ovat moninaiset, ja tyypillisesti tutkimuksessa yhdistetään monenlaisia aineistoja ja useampia menetelmiä (ks. Lingvistinen etnografia toimintayhteisöissä tk.).

Sekä haastattelujen että kyselyjen avulla on selvitetty kääntäjien ja tulkkien näkemyksiä heidän ammatillisesta statuksestaan ja identiteetistään sekä roolistaan. Esimerkiksi Katan (2009) teki maailmanlaajuisen internetkyselyn, jossa selvitettiin vastaajien näkemyksiä kääntäjän ja tulkin roolista, ammatin statuksesta, ammatillisesta tyytyväisyydestä ja koulutuksesta. Kyselyyn vastasi noin 1 000 kääntäjää ja tulkkia sekä kääntämisen opettajaa ja opiskelijaa. Damin ja Zethsenin kyselytutkimukset (esim. 2012) eri tehtävissä toimivien tanskalaisten kääntäjien statuksesta ovat olleet esimerkkeinä myös suomalaisille vastaaville tutkimuksille (esim. Ruokonen 2016).

Etnografinen tutkimus on tyypillisesti monimenetelmäistä, ja se pyrkii antamaan kohteestaan monipuolisen ja rikkaan kuvauksen. Antropologian käsite *tiheä kuvaus* (engl. *thick description*) on siirretty käännöstieteeseen käsitteellä *thick translation* (Appiah 1993). Tyypillisesti tällainen etnografinen tutkimus yhdistää havainnointia, haastattelua tai kyselyä sekä tekstien tai dokumenttien analyysia. Varhaisimpia laajoja etnografisia tutkimuksia on Susan Berk-Seligsonin (1990) tutkimus espanja–englanti-oikeustulkkauksesta Yhdysvalloissa. Hän havainnoi oikeussalissa tulkkausta päivittäin seitsemän kuukauden ajan, nauhoitti tulkkauksia yhteensä 114 tuntia ja analysoi nauhoitukset. Lisäksi hän haastatteli sekä tulkkeja että viranomaisia. Kaisa Koskisen (2008) etnografinen tutkimus tarkasteli EU:n komission kääntämistä. Hän tutki institutionaalista kontekstia ja kääntäjiä, jotka toimivat tässä kontekstissa, sekä käännettyjä dokumentteja ja niiden lähdetekstejä. Menetelminä olivat fokusryhmähaastattelu, kysely, (osallistuva) havainnointi ja tekstien analyysi (käännössiirtymien analyysi ja kriittinen diskurssianalyysi). Koskisen tutkimus on myös esimerkki käännöstieteellisestä tapaustutkimuksesta; tapaus on rajattu Euroopan komission Suomen Luxemburgissa sijaitsevaan käännösyksikköön. Tapaustutkimus on niin ikään Hokkasen (2016) autoetnografinen tutkimus, jossa hän havainnoi

tulkkausta helluntaiseurakunnassa, piti kirjaa omista tulkkauskokemuksistaan ja nauhoitti omia tulkkauksiaan.

Kuten edelliset esimerkit osoittavat, etnografisissa tutkimuksissa yhdistyvät usein tuotteen, prosessin, toimijoiden, toimintaympäristön ja vastaanoton tutkimukset, jotka tässä artikkelissa olemme selvyyden vuoksi osin erottaneet omiksi luvuikseen. Esimerkiksi Ebru Dirikerin (2004) tutkimuksessa konferenssitulkkauksesta aineistona oli nauhoitettuja tulkkauksia, tulkkauksesta käytyä julkista keskustelua sekä konferenssitulkkien, järjestäjien, puhujien ja osallistujien haastatteluja. Kuten Dirikerin tutkimuksessa käännöstieteellinen tutkimus on kääntäjien ja tulkkien lisäksi kiinnostunut myös muista kääntämisen ja tulkkauksen kentän toimijoista.

Etnografisessa tutkimuksessa käytetään myös bibliografisia aineistoja ja arkistodokumentteja; kiinnostuksen kohteena voi olla esimerkiksi tietyn aikakauden käännös- ja tulkkaustoiminta (Kujamäki & Paloposki 2015). Historiallinen tutkimus voi kohdistua myös yhteen henkilöön, esimerkkinä Paloposken tutkimus Samuli Suomalaisesta (1850–1907) kääntäjänä (Paloposki 2017).

Kenttätutkimuksesta uudenlaisia esimerkkejä ovat translationaalisen tilan tutkimus eli monikielisen tilan tarkasteleminen kääntämisen paikkana (Koskinen 2013a), kuten Tampereen kielimaiseman tutkimus ”käännettyyden” näkökulmasta (Koskinen 2013b) ja Helsinki–Pietari-väliä pendelöivän Allegro-junan monikielisen viestinnän tutkimus (Probirskaja 2017). Molemmissa tutkimuksissa havainnointi on keskeinen aineiston keruun menetelmä, edellisessä aineistona on lisäksi valokuvia, jälkimmäisessä ohjeita ja haastatteluja.

Kääntämisen ja tulkkauksen vastaanoton tutkimuksessa tarkastellaan käännöksen tai tulkkauksen käytettävyyttä, hyväksyttävyyttä ja ymmärrettävyyttä, yleisesti ottaen laatua. Laatu on riippuvainen kääntämisen tai tulkkauksen tarkoituksesta ja vastaanottajan tarpeista. Esimerkiksi konferenssitulkkauksen osanottajilla, oikeusprosessin viranomaisilla ja lääkärin vastaanoton potilaalla voi olla hyvin erilaisia odotuksia tulkkaukselle, minkä vuoksi tulkkauksen laadulle on vaikea määritellä yleisiä kriteereitä. Odotukset voivat vaihdella myös sen mukaan, mitä kussakin kulttuurissa minäkin aikana pidetään hyvänä tulkkauksena.

Esimerkiksi monessa maassa juristit odottavat oikeustulkkaukselta sanasanaista tulkkausta (Kalina 2016), mutta Kadricin (2001) kyselytutkimuksen mukaan Itävallassa tuomarit pitävät tärkeämpänä tyylin ja rekisterin uskollisuutta. Konferenssitulkkausta koskevissa kyselytutkimuksissa (esim. Vuorikoski 1996) on yleisesti todettu, että keskeisiä laatutekijöitä ovat uskollisuus alkuperäiselle puheelle, sisältö, synkronia, retoriset taidot ja äänenlaatu. Tärkeitä ovat niin ikään täydellisyys, ilmauksen selkeys ja terminologian tarkkuus. Toisaalta on myös havaittu, että odotukset vaihtelevat muun muassa kokoustyypin ja osallistujien simultaanitulkkauskokemuksen mukaan. (Pöchhacker 2004, 154.) Vastaavia kyselytutkimuksia on tehty myös esimerkiksi asiatekstien käännösten tilaajille (esim. Havumetsä 2012).

Vastaanoton tutkimuksessa voidaan käyttää autenttisten käännösten ja tulkkausten lisäksi koeasetelmia, joissa tutkitaan, miten käännöksen tai tulkkauksen eri tekijät vaikuttavat vastaanottoon. Esimerkiksi Hale, Martschuk, Ozolins ja Stern (2017) ovat tutkineet simuloitujen oikeustulkkausten avulla sitä, miten tulkkauslaji (konsekutiivi vs. simultaani) vaikuttaa lautamiesten arviointiin todistajan uskottavuudesta.

Uusista menetelmistä silmänliiketutkimus soveltuu myös tekstin luettavuuden ja ymmärrettävyyden tutkimukseen; esimerkiksi O'Brien (2010) on tutkinut konekäännetyn tekstin luettavuutta. Taivalkoski-Shilov ja Koponen (2017) ovat yhdistäneet silmänliiketutkimusta, haastatteluja ja osallistujien narratiivisia raportteja tutkiessaan, miten paratekstit ja lukijoiden akateeminen tausta vaikuttavat tieteellisen tekstin vastaanottoon. Silmänliiketutkimusta on käytetty lisäksi erityisesti tekstityksen vastaanottotutkimuksissa, joissa on selvitetty muun muassa tekstin määrän ja muodon (tiivistetty vs. puhetta tarkkaan noudattava) vaikutusta ymmärrettävyyteen ja elokuvan katsomiseen (esim. Szarkowska, Krejtz, Klyszejko & Wieczorek 2011). Lång, Mäkisalo, Gowases ja Pietinen (2013) selvittivät silmänliiketutkimuksen avulla sitä, miten ruututekstityksen konventioiden noudattaminen vaikuttaa vastaanottoon. Silmänliiketutkimukseen yhdistetään usein kysely, jolla testataan tekstitetyn ohjelman ymmärtämistä.

Tekstityksen vastaanottoa on tutkittu myös fokusryhmien avulla, jolloin kohderyhmään kuuluvat vastaanottajat voivat keskustella

tekstityksestä sekä katsoessaan tekstitettyä elokuvaa että välittömästi katsomisen jälkeen (esim. Tuominen 2012; Tiittula 2016). Fokusryhmätutkimuksen etuna esimerkiksi kyselyihin verrattuna on se, että yleisten näkemysten lisäksi saadaan konkreettisiin esiintymiin kommentteja sekä argumentointia eri ratkaisujen puolesta ja niitä vastaan. Usein vastaanottotutkimuksissakin yhdistetään monia menetelmiä, kuten Orrego-Carmona (2016) teki tutkiessaan, erottavatko katsojat ammattitekstittäjän tekstityksen harrastajatekstittäjän tuotoksesta. Tutkimuksessa katsojille näytettiin eri tavoin tekstitettyjä elokuvakatkelmia ja vastaanottajien reaktioita tutkittiin silmänliikkeiden, kyselyn ja haastattelun avulla. Käännetyn kirjallisuuden vastaanoton tutkimuksessa voidaan käyttää edellä mainittujen menetelmien lisäksi arvosteluja, julkisia keskusteluja, lukijapalautteita sekä tilastoja.

**Toimijoiden ja toimintaympäristön tutkimusta:**

Angelelli, Claudia (toim.). 2014. *The Sociological Turn in Translation and Interpreting Studies*. Amsterdam: John Benjamins.

Biagin, Marta, Boyd, Michael S. & Monacelli, Claudia (toim.). 2017. *The Changing Role of the Interpreter. Contextualising Norms, Ethics and Quality Standards*. New York (NY): Routledge.

Dam, Helle V. & Zethsen, Karen Korning (toim.). 2009. Translation studies: Focus on the translator. *Hermes*, Special issue 42. Saatavissa: https://tidsskrift.dk/her/issue/view/2853.

Kinnunen, Tuija & Koskinen, Kaisa (toim.). 2010. *Translators' Agency*. Tampere Studies in Language, Translation and Culture. Series B 4. [verkkoaineisto]. Tampere: Tampere University Press. Saatavissa: http://urn.fi/urn:isbn:978-951-44-8082-9.

Sela-Sheffy, Rakefet & Shlesinger, Miriam (toim.). 2011. *Identity and Status in the Translational Professions*. Amsterdam: John Benjamins.

Taivalkoski-Shilov, Kristiina, Tiittula, Liisa & Koponen, Maarit (toim.). 2017. *Communities in Translation and Interpreting*. Vita traductiva 9. Montréal: Éditions québécoises de l'œuvre.

**Kääntämisen historian tutkimuksen menetelmistä:**

Pym, Anthony. 1998. *Method in Translation History*. Manchester: St. Jerome.

**Käännöstieteellisistä vastaanottotutkimuksista ja niissä käytetyistä menetelmistä:**

Suojanen, Tytti, Koskinen, Kaisa & Tuominen, Tiina. 2012. *Käyttäjäkeskeinen kääntäminen*. [verkkoaineisto]. Tampere: Tampereen yliopisto. Saatavissa: http://urn.fi/URN:ISBN:978-951-44-8839-9.

# 4. Yhteenveto

Katsaus käännöstieteen historiaan, tutkimusaineistoihin, -menetelmiin ja -kysymyksiin nostaa esille tieteenalan itsenäistymisen emotieteenaloista. Temaattisesti kehitys on kulkenut produktin, kirjoitettujen käännösten ja puhuttujen tulkkeiden tutkimuksesta multimodaalisiin tuotoksiin ja vuorovaikutustilanteisiin sekä toimintaprosessien, toimijoiden, toimintaympäristöjen ja vastaanoton kysymysten tutkimukseen. Menetelmällisesti kehitys on yhtäältä säilyttänyt perinteisiä tekstianalyysin manuaalisia menetelmiä mutta tuonut toisaalta mukanaan siirtymän tallennuksessa, mittauksessa ja analyysissa tietokoneohjelmiin sekä sähköisiin mittaus- ja tallennusvälineisiin. Mukaan on tullut myös toimintakentän, toimijoiden ja vastaanottajien etnografinen tutkimus. Yhä useammin tutkimus on moniaineistoista ja monimenetelmäistä. Menetelmien ja välineiden kehittyessä on kiinnitetty myös enemmän huomiota tutkimuskysymyksiin, jotka nousevat yhteiskunnasta. Kääntämisen ja tulkkauksen ilmiöt ovat jatkuvan muutoksen alaisia. Yhteiskunnallisten ja teknologisten muutosten myötä syntyy uusia kääntämisen ja tulkkauksen alueita ja muotoja, joita on vaikea sijoittaa vanhoihin kategorioihin. Tällainen on esimerkiksi reaaliaikainen tekstitys. Käännös- ja kieliteknologia, ihmisen ja koneen

välinen vuorovaikutus sekä yhä monikielisempi yhteiskunta ja saavutettavuuden haasteet asettavat uusia tehtäviä käännöstieteelliselle tutkimukselle. Samalla vaatimus tieteidenvälisyydestä on kasvanut, vaikka käännöstiede onkin aina ollut pohjimmiltaan monitieteistä.

**Aiheesta lisää:**

**Käännöstieteellisen tutkimuksen ajankohtaisia käsikirjoja:**

Gambier, Yves & Doorslaer, Luc van (toim.). 2010–2016. *Handbook of Translation Studies Online*. [verkkoaineisto]. Amsterdam: John Benjamins.

Malmkjær, Kirsten & Windle, Kevin (toim.). 2011. *Oxford Handbook of Translation Studies*. Oxford: Oxford University Press.

Millán, Carmen & Bartrina, Francesca (toim.). 2013. *Routledge Handbook of Translation Studies*. London: Routledge.

Pöchhacker, Franz (toim.). 2015. *Routledge Encyclopedia of Interpreting Studies*. New York (NY): Routledge.

**Käännöstieteellisen tutkimuksen ajankohtaisia metodioppaita:**

Angelelli, Claudia & Baer, Brian James (toim.). 2015. *Researching Translation and Interpreting*. New York (NY): Routledge.

Bogucki, Łukasz. 2016. *Areas and Methods of Audiovisual Translation Research*. (2. uud. laitos). Frankfurt am Main: Peter Lang Edition.

Hale, Sandra & Napier, Jemina. 2013. *Research Methods in Interpreting: A Practical Resource*. New York (NY): Continuum.

Olohan, Maeve (toim.). 2000. *Intercultural Faultlines. Research Modals in Translation Studies I. Textual and Cognitive Aspects*. Manchester: St. Jerome.

Saldanha, Gabriela & O'Brien, Sharon. 2013. *Research Methodologies in Translation Studies.* London: Routledge.

**Muita hyödyllisiä lähteitä:**

Gambier, Yves & Doorslaer, Luc van (toim.). 2004–. *Translation Studies Bibliography*. [verkkoaineisto]. Amsterdam: John Benjamins.

## VIITTEET

1 http://www.hf.uio.no/ilos/english/services/omc/enpc/.
2 https://www.alc.manchester.ac.uk/translation-and-intercultural-studies/research/projects/translational-english-corpus-tec/.
3 http://metashare.csc.fi/repository/browse/the-finnish-broadcasting-company-corpus-of-subtitles/6c23ccd8fccc11e18b49005056be118ed4f163dbb7344fc3bcc-4969c9500721e/.
4 http://opus.lingfil.uu.se/.
5 Katsauksen kääntämisen universaalien tutkimukseen luo esimerkiksi Mauranen (2006).
6 http://metashare.elda.org/repository/browse/european-parliament-interpretation-corpus-epic/fde33884de7611e2b1e400259011f6ea48ac8ce0b41f48e6be224bbee-9d59cb9/.
7 http://www.yorku.ca/comindat/comindat.htm.

## KIRJALLISUUS


Abdallah, Kristiina. 2012. *Translators in Production Networks. Reflections on Agency, Quality and Ethics*. Publications of the University of Eastern Finland. Dissertations in Education, Humanities, and Theology 21. Joensuu: Itä-Suomen yliopisto. [verkkoaineisto]. [viitattu 23.12.2017]. Saatavissa: http://urn.fi/URN:ISBN:978-952-61-0609-0.

Angermeyer, Philipp Sebastian, Meyer, Bernd & Schmidt, Thomas. 2012. Sharing community interpreting corpora: a pilot study. Julkaisussa: Schmidt, Thomas & Wörner, Kai (toim.) *Multilingual Corpora and Multilingual Corpus Analysis*. Amsterdam: John Benjamins, 275–294.

Appiah, Kwame Anthony. 1993. Thick translation. *Callaloo* 16:4, 808–819.

Baker, Mona. 1993. Corpus linguistics and translation studies. Implications and applications. Julkaisussa: Baker, Mona, Francis, Gill & Tognini Bonelli, Elena (toim.) *Text and Technology. In Honour of John Sinclair*. Amsterdam: John Benjamins, 233–250.

Baker, Mona. 1996. Corpus-based translation studies: The challenges that lie ahead. Julkaisussa: Somers, Harold (toim.) *Terminology, LSP and Translation. Studies in Language Engineering in Honour of Juan C. Sager*. Amsterdam: John Benjamins, 175–186.

Becher, Viktor. 2011. *Explicitation and Implicitation in Translation. A Corpus-based Study of English–German and German–English Translations of Business Texts*. Väitöskirja. Hampuri:

Hampurin yliopisto. [verkkoaineisto]. [viitattu 14.1.2018]. Saatavissa: https://d-nb.info/102042673X/34.
Bendazzoli, Claudio. 2015. Corpus-based research. Julkaisussa: Pöchhacker, Franz (toim.) *Routledge Encyclopedia of Interpreting Studies*. Florence: Taylor & Francis, 132–135.
Bendazzoli, Claudio. 2018. Corpus-based interpreting studies: past, present and future developments of a (wired) cottage industry. Julkaisussa: Russo, Mariachiara, Bendazzoli, Claudio & Defrancq, Bart (toim.) *Making Way in Corpus-based Interpreting Studies*. New York (NY): Springer, 1–20.
Berk-Seligson, Susan. 1990. *The Bilingual Courtroom: Court Interpreters in the Judicial Process*. Chicago (IL): The University of Chicago Press.
Bernardini, Silvia, Ferraresi, Adriano, Russo, Mariachiara, Collard, Camille & Defrancq, Bart. 2018. Building interpreting and intermodal corpora: A how-to for a formidable task. Julkaisussa: Russo, Mariachiara, Bendazzoli, Claudio & Defrancq, Bart (toim.) *Making Way in Corpus-based Interpreting Studies*. New York (NY): Springer, 21–42.
Braun, Sabine & Taylor, Judith L. 2011. AVIDICUS comparative studies – part I: Traditional interpreting and remote interpreting in police interviews. Julkaisussa: Braun, Sabine & Taylor, Judith L. (toim.) *Videoconference and Remote Interpreting in Criminal Proceedings*. Guildford: University of Surrey, 85–100.
Bundgaard, Kristine, Paulsen Christensen, Tina & Schjoldager, Anne. 2016. Translator-computer interaction in action – an observational process study of computer-aided translation. *The Journal of Specialised Translation* 25, 106–130. [verkkoaineisto]. [viitattu 29.12.2017]. Saatavissa: http://www.jostrans.org/issue25/art_bundgaard.pdf.
Cappelle, Bert & Loock, Rudy. 2017. Typological differences shining through. The case of phrasal verbs in translated English. Julkaisussa: Delaere, Isabelle, Lefer, Marie-Aude & De Sutter, Gert (toim.) *Empirical Translation Studies: New Methodological and Theoretical Traditions*. Berlin: De Gruyter Mouton, 235–264.
Castagnoli, Sara, Ciobanu, Dragoş, Kunz, Kerstin, Kübler, Natalie & Volanschi, Alexandra. 2011. Designing a translator learner corpus for training purposes. Julkaisussa: Kübler, Natalie (toim.) *Corpora, Language, Teaching and Resources: from Theory to Practice*. Bern: Lang, 221–248. [verkkoaineisto]. [viitattu 17.10.2017]. Saatavissa: http://www.eila.univ-paris-diderot.fr/_media/user/alexandra_volanschi/publi/castagnoli_et_al.pdf.
Catford, John Cunnison. 1965. *A Linguistic Theory of Translation: An Essay in Applied Linguistics*. London: Oxford University Press.
Chesterman, Andrew. 2004. Beyond the particular. Julkaisussa: Mauranen, Anna & Kujamäki, Pekka (toim.) *Translation Universals: Do they Exist?* Amsterdam: John Benjamins, 33–49.
Chesterman, Andrew. 2016 [1997]. *Memes of Translation: The Spread of Ideas in Translation Theory*. Amsterdam: John Benjamins.
Dam, Helle V. & Zethsen, Karen Korning. 2012. Translators in international organizations. A special breed of high-status professionals? Danish EU translators as a case in point. *Translation and Interpreting Studies* 7:2, 211–232.
Delaere, Isabelle & De Sutter, Gert. 2017. Variability of English loanword use in Belgian Dutch translations. Measuring the effect of source language and register. Julkaisussa: Delaere, Isabelle, Lefer, Marie-Aude & De Sutter, Gert (toim.) *Empirical Translation*

*Studies: New Methodological and Theoretical Traditions*. Berlin: De Gruyter Mouton, 81–112.
Delaere, Isabelle, Lefer, Marie-Aude & De Sutter, Gert (toim.). 2017. *Empirical Translation Studies: New Methodological and Theoretical Traditions*. Berlin: De Gruyter Mouton.
Diriker, Ebru. 2004. *De-/re-contextualizing Conference Interpreting: Interpreters in the Ivory Tower?* Amsterdam: John Benjamins.
Dragsted, Barbara. 2010. Coordination of reading and writing processes in translation: an eye on uncharted territory. Julkaisussa: Shreve, Gregory M. & Angelone, Erik (toim.) *Translation and Cognition*. Amsterdam: John Benjamins, 41–62.
Englund Dimitrova, Birgitta. 2005. *Expertise and Explicitation in the Translation Process*. Amsterdam: John Benjamins.
Englund Dimitrova, Birgitta. 2016. Translation process. Julkaisussa: Gambier, Yves & Doorslaer, Luc van (toim.) 2010–2016. *Handbook of Translation Studies Online*. Amsterdam: John Benjamins. [verkkoaineisto]. [viitattu 30.12.2017]. Saatavissa: DOI: 10.1075/hts.1.tra6.
Englund Dimitrova, Birgitta & Hyltenstam, Kenneth (toim.). 2000. *Language Processing and Simultaneous Interpreting: Interdisciplinary Perspectives*. Amsterdam: John Benjamins.
Even-Zohar, Itamar. 1990. *Polysystem Studies*. Durham: Duke University Press. Special issue of Poetics Today 11:1.
Evert, Stefan & Neumann, Stella. 2017. The impact of translation direction on characteristics of translated texts. A multivariate analysis for English and German. Julkaisussa: Delaere, Isabelle, Lefer, Marie-Aude & De Sutter, Gert (toim.) *Empirical Translation Studies: New Methodological and Theoretical Traditions*. Berlin: De Gruyter Mouton, 47–80.
Fetterman, David M. 2011. Ethnography. Julkaisussa: Lewis-Beck, Michael S., Bryman, Alan & Liao, Tim Futing (toim.) *The SAGE Encyclopedia of Social Science Research Methods*. Thousand Oaks (CA): Sage, 329–332.
Gavioli, Laura & Baraldi, Claudio. 2011. Interpreter-mediated interaction in healthcare and legal settings: talk organization, context and the achievement of intercultural communication. *Interpreting* 13:2, 205–233.
Gile, Daniel. 1997. Conference interpreting as a cognitive management problem. Julkaisussa: Danks, Joseph H., Shreve, Gregory M., Fountain, Stephen B., & McBeath, Michael K. (toim.) *Cognitive Processes in Translation and Interpreting*. Thousand Oaks (CA): Sage, 196–214.
Gile, Daniel. 2009. Interpreting studies: A critical view from within. *MonTI: Monografías de traducción e interpretación* 1, 135–155.
Gile, Daniel. 2015. Experimental research. Julkaisussa: Angelelli, Claudia & Baer, Brian James (toim.) *Researching Translation and Interpreting*. New York (NY): Routledge, 220–228.
Göpferich, Susanne, Bayer-Hohenwarter, Gerrit, Prassl, Friederike & Stadlober, Johanna. 2011. Exploring translation competence acquisition: criteria of analysis put to the test. Julkaisussa: O'Brien, Sharon (toim.) *Cognitive Explorations of Translation*. London: Continuum, 57–85.
Hale, Sandra, Martschuk, Natalie, Ozolins, Uldis & Stern, Ludmila. 2017. The effect of interpreting modes on witness credibility assessments. *Interpreting* 19:1, 69–96.

Hale, Sandra & Napier, Jemina 2013. *Research Methods in Interpreting: A Practical Resource*. New York (NY): Continuum.

Hansen, Gyde. 2006. Retrospection methods in translator training and translation research. *The Journal of Specialised Translation* 5, 2–41. [verkkoaineisto]. [viitattu 29.12.2017]. Saatavissa: http://www.jostrans.org/issue05/art_hansen.php.

Harjunpää, Katariina. 2017. *Translatory Practices in Everyday Conversation. Bilingual Mediating in Finnish-Portuguese Interaction*. Helsinki: Helsingin yliopisto. [verkkoaineisto]. [viitattu 21.8.2017]. Saatavissa: http://urn.fi/URN:ISBN: 978-951-51-3227-7.

Havumetsä, Nina. 2012. *The Client Factor: A Study of Clients' Expectations Regarding Non-literary Translators and the Quality of Non-literary Translations*. Helsinki: Helsingin yliopisto. [verkkoaineisto]. [viitattu 19.12.2017]. Saatavissa: http://urn.fi/URN:ISBN:978-952-10-8392-1.

Hirvonen, Maija & Tiittula, Liisa. 2018. How are translations created? Using multimodal conversation analysis to study a team translation process. *Linguistica Antverpiensia, New Series: Themes in Translation Studies* 17, 157–173. [verkkoaineisto]. [viitattu 24.2.2019]. Saatavissa: https://lans-tts.uantwerpen.be/index.php/LANS-TTS/issue/view/19.

Hokkanen, Sari. 2016. *To Serve and to Experience: An Autoethnographic Study of Simultaneous Church Interpreting*. Tampere: Tampereen yliopisto. [verkkoaineisto]. [viitattu 19.12.2017]. Saatavissa: http://urn.fi/URN:ISBN:978-952-03-0232-0.

Holmes, James S. 2000 [1972]. The name and nature of translation studies. Julkaisussa: Venuti, Lawrence (toim.) *The Translation Studies Reader*. London: Routledge, 172–191.

Holz-Mänttäri, Justa. 1984. *Translatorisches Handeln: Theorie und Methode*. Helsinki: Suomalainen Tiedeakatemia.

Isolahti, Nina. 2014. *Tulkkauksen tarkkuus rikosoikeudenkäynnissä – saavuttamaton ihanne*. Acta Electronica Universitatis Tamperensis 1374. Tampere: Tampereen yliopisto. [verkkoaineisto]. [viitattu 23.12.2017]. Saatavissa: https://tampub.uta.fi/bitstream/handle/10024/94867/978-951-44-9329-4.pdf?sequence=1.

Jakobsen, Arnt Lykke. 2003. Effects of think aloud on translation speed, revision and segmentation. Julkaisussa: Alves, Fabio (toim.) *Triangulating Translation. Perspectives in Process Oriented Research*. Amsterdam: John Benjamins, 69–95.

Jakobsen, Arnt Lykke & Jensen, Kristian T. H. 2008. Eye movement behaviour across four different types of reading task. Julkaisussa: Mees, Inger M., Jacobsen, Arnt Lykke & Göpferich, Susanne (toim.) *Looking at Eyes: Eye-tracking Studies of Reading and Translation Processing*. Frederiksberg: Samfundslitteratur Press, 103–124.

Jakobson, Roman. 2000 [1959]. On linguistic aspects of translation. Julkaisussa: Venuti, Lawrence (toim.) *The Translation Studies Reader.* London: Routledge, 113–118.

Jiménez-Crespo, Miguel A. 2016. Translation crowdsourcing: research trends and perspectives. Julkaisussa: Cordingley, Anthony & Frigau Manning, Céline (toim.) *Collaborative Translation: From the Renaissance to the Digital Age*. London: Bloomsbury, 192–211.

Johansson, Stig. 2007. *Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies*. Amsterdam: John Benjamins.

Jääskeläinen, Riitta. 1999. *Tapping the Process: An Explorative Study of the Cognitive and Affective Factors Involved in Translating*. Väitöskirja. Joensuu: Joensuun yliopisto.

Jääskeläinen, Riitta. 2011b. Back to basics: designing a study to determine the validity and reliability of verbal report data on translation process. Julkaisussa: O'Brien, Sharon (toim.) *Cognitive Explorations of Translation*. London: Continuum, 15–29.

Kadric, Mira. 2001. *Dolmetschen bei Gericht: Erwartungen, Anforderungen, Kompetenzen*. Wien: WUV Universitätsverlag.

Kalina, Sylvia. 2016. Quality in interpreting. Julkaisussa: Gambier, Yves & Doorslaer, Luc van (toim.) *Handbook of Translation Studies Online*. Amsterdam: John Benjamins. [verkkoaineisto]. [viitattu 18.12.2017]. Saatavissa: DOI: 10.1075/hts.3.qua2.

Katan, David. 2009. Occupation or profession: a survey of the translators' world. *Translation and Interpreting Studies* 4:2, 187–209.

Kemppanen, Hannu. 2008. *Avainsanoja ja ideologiaa: käännettyjen ja ei-käännettyjen historiatekstien korpuslingvistinen analyysi*. Väitöskirja. Joensuu: Joensuun yliopisto.

Kemppanen, Hannu. 2011. Pamphlet or scholarly work. Book reviews and determining the place of a translation. Julkaisussa: Kujamäki, Pekka, Kolehmainen, Leena, Penttilä, Esa & Kemppanen, Hannu (toim.) *Beyond Borders – Translations Moving Languages, Literatures and Cultures*. Berlin: Frank & Timme, 145–162.

Kinnunen, Tuija. 2010. Agency, activity and court interpreting. Julkaisussa: Kinnunen, Tuija & Koskinen, Kaisa (toim.) *Translators' Agency*. Tampere Studies in Language, Translation and Culture. Series B 4. Tampere: Tampere University Press, 126–164. [verkkoaineisto]. [viitattu 23.12.2017]. Saatavissa: http://urn.fi/urn:isbn:978-951-44-8082-9.

Kinnunen, Tuija. 2013. Oikeusturva, oikeustulkkaus ja translationaali tila. Julkaisussa: Koskinen, Kaisa (toim.) *Tulkattu Tampere*. Tampere: Tampere University Press, 71–105.

Kinnunen, Tuija & Koskinen, Kaisa (toim.). 2010. *Translators' Agency*. Tampere Studies in Language, Translation and Culture. Series B 4. Tampere: Tampere University Press. [verkkoaineisto]. [viitattu 23.12.2017]. Saatavissa: http://urn.fi/urn:isbn:978-951-44-8082-9.

Koglin, Arlene. 2015. An empirical investigation of cognitive effort required to post-edit machine translated metaphors compared to the translation of metaphors. *Translation & Interpreting* 7:1, 126–141.

Kolehmainen, Leena, Miestamo, Matti & Nordlund, Taru (toim.). 2013. *Kielten vertailun metodiikka*. Suomalaisen Kirjallisuuden Seuran Toimituksia 1387. Helsinki: Suomalaisen Kirjallisuuden Seura.

Koller, Werner. 1979. *Einführung in die Übersetzungswissenschaft*. Heidelberg: Quelle & Meyer.

Koskela, Merja, Koskinen, Kaisa & Pilke, Nina. 2017. Bilingual formal meeting as a context of translatoriality. *Target* 29:3, 464–485.

Koskinen, Kaisa. 2008. *Translating Institutions: An Ethnographic Study of EU Translation*. Manchester: St. Jerome.

Koskinen, Kaisa (toim.). 2013a. *Tulkattu Tampere*. Tampere: Tampere University Press.

Koskinen, Kaisa. 2013b. Turistina Hervannan kielimaisemassa: käännettyyden jäljillä. Julkaisussa: Koskinen, Kaisa (toim.) *Tulkattu Tampere*. Tampere: Tampere University Press, 313–334.

Koskinen, Kaisa & Paloposki, Outi. 2015. *Sata kirjaa, tuhat suomennosta: kaunokirjallisuuden uudelleenkääntäminen*. Suomalaisen Kirjallisuuden Seuran Toimituksia 1410. Helsinki: Suomalaisen Kirjallisuuden Seura.
Kruger, Haidee. 2017. The effects of editorial intervention. Implications for studies of the features of translated language. Julkaisussa: Delaere, Isabelle, Lefer, Marie-Aude & De Sutter, Gert (toim.) *Empirical Translation Studies: New Methodological and Theoretical Traditions*. Berlin: De Gruyter Mouton, 113–156.
Kruger, Jan-Louis. 2012. Making meaning in AVT: Eye tracking and viewer construction of narrative. *Perspectives: Studies in Translatology* 20:1, 67–86.
Kujamäki, Pekka. 1998. *Deutsche Stimmen der Sieben Brüder. Ideologie, Poetik und Funktionen literarischer Übersetzung*. Frankfurt a.M.: Peter Lang.
Kujamäki, Pekka & Paloposki, Outi. 2015. Sujuvasti suomesta saksaan kääntävä konekirjoittajatar saa paikan heti: historian näkökulma kääntäjyyteen ammattina tai tehtävänä. *Virittäjä* 119:3, 342–371.
Kumpulainen, Minna. 2016. *Learning Translation: An Empirical Study into the Acquisition of Interlingual Text Production Skills*. Publications of the University of Eastern Finland. Dissertations in Education, Humanities, and Theology, no 91. Joensuu: Itä-Suomen yliopisto. [verkkoaineisto]. [viitattu 15.12.2017]. Saatavissa: http://urn.fi/URN:ISBN:978-952-61-2235-9.
Kussmaul, Paul. 1995. *Training the Translator*. Philadelphia: John Benjamins.
Kuusi, Päivi. 2011. *Miksi näkökulma muuttuu käännöksessä? Eksplisiittistämisen ja normaalistamisen selitysvoima ja seuraukset*. Väitöskirja. Tampere: Tampereen yliopisto. [verkkoaineisto]. [viitattu 23.12.2017]. Saatavissa: http://tampub.uta.fi/bitstream/handle/10024/66777/978-951-44-8522-0.pdf?sequence=1&isAllowed=y.
Kuusi, Päivi. 2016. Features of discourse presentation in translation: Literary and narratological insights into translation universals. *International Journal of Literary Linguistics* 5:3, 1–24.
Laviosa, Sara. 1996. *The English Comparable Corpus (ECC): A Resource and Methodology for the Empirical Study of Translation*. Manchester: University of Manchester.
Laviosa, Sara, Pagano, Adriana, Kemppanen, Hannu & Ji, Meng. 2017. *Textual and Contextual Analysis in Empirical Translation Studies*. Singapore: Springer.
Lång, Juha, Mäkisalo, Jukka, Gowases, Tersia & Pietinen, Sami. 2013. Using eye tracking to study the effect of badly synchronized subtitles on the gaze paths of television viewers. *New Voices in Translation Studies* 10, 72–86. [verkkoaineisto]. [viitattu 29.12.2017]. Saatavissa: https://www.iatis.org/index.php/new-voices-in-translation-studies/item/794-issue10-2013.
Mason, Ian. 2012. Gaze, positioning and identity in interpreter-mediated dialogues. Julkaisussa: Baraldi, Claudio & Gavioli, Laura (toim.) *Coordinating Participation in Dialogue Interpreting*. Amsterdam: John Benjamins, 177–200.
Mauranen, Anna. 2000. Strange strings in translated language. A study on corpora. Julkaisussa: Olohan, Maeve (toim.) *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects*. Manchester: St. Jerome, 119–141.

Mauranen, Anna. 2004. Corpora, universals and interference. Julkaisussa: Mauranen, Anna & Kujamäki, Pekka (toim.) *Translation Universals. Do They Exist?* Amsterdam: John Benjamins, 65–82.

Mauranen, Anna. 2006. Translation universals. Julkaisussa: Brown, Keith, Anderson, Anne, Bauer, Laurie, Berns, Margie, Hirst, Graeme & Miller, Jim (toim.) *Encyclopedia of Language and Linguistics*. Amsterdam: Elsevier, 93–100.

Monti, Cristina, Bendazzoli, Claudio, Sandrelli, Annalisa & Russo, Mariachiara. 2005. Studying directionality in simultaneous interpreting through an electronic corpus: EPIC (European Parliament Interpreting Corpus). *Meta* 50:4. [verkkoaineisto]. [viitattu 17.10.2017]. Saatavissa: https://www.erudit.org/en/journals/meta/2005-v50-n4-meta1024/019850ar.pdf.

Munday, Jeremy. 2012. *Introducing Translation Studies: Theories and Applications*. (3. p.). New York (NY): Routledge.

Munday, Jeremy. 2015. The role of archival and manuscript research in the investigation of translator decision-making. Julkaisussa: Ehrensberger-Dow, Maureen, Göpferich, Susanne & O'Brien, Sharon (toim.) *Interdisciplinarity in Translation and Interpreting Process Research*. Amsterdam: John Benjamins, 127–140.

Muñoz Martín, Ricardo. 2010. On paradigms and cognitive translatology. Julkaisussa: Shreve, Gregory M. & Angelone, Erik (toim.) *Translation and Cognition*. Amsterdam: John Benjamins, 169–187.

Nida, Eugene Albert. 1964. *Toward a Science of Translating: With Special Reference to Principles and Procedures Involved in Bible Translating*. Leiden: Brill.

O'Brien, Sharon. 2006. Pauses as indicators of cognitive effort in post-editing machine translation output. *Across Languages and Cultures* 7:1, 1–21.

O'Brien, Sharon. 2010. Controlled language and readability. Julkaisussa: Shreve, Gregory M. & Angelone, Erik (toim.) *Translation and Cognition*. Amsterdam: John Benjamins, 143–165.

O'Brien, Sharon (toim.). 2011. *Cognitive Explorations of Translation*. London: Continuum.

Orrego-Carmona, David. 2016. A reception study on non-professional subtitling: do audiences notice any difference? *Across Languages and Cultures* 17:2, 163–181.

PACTE Group. 2011. Results of the validation of the PACTE translation competence model: translation project and dynamic translation index. Julkaisussa: O'Brien, Sharon (toim.) *Cognitive Explorations of Translation*. London: Continuum, 30–56.

Paloposki, Outi. 2007. Suomentaminen ja suomennokset 1800-luvulla. Julkaisussa: Riikonen, H. K., Kovala, Urpo, Kujamäki, Pekka & Paloposki, Outi (toim.) *Suomennoskirjallisuuden historia I*. Suomalaisen Kirjallisuuden Seuran Toimituksia 1084. Helsinki: Suomalaisen Kirjallisuuden Seura, 102–126.

Paloposki, Outi. 2017. In search of an ordinary translator: translator histories, working practices and translator–publisher relations in the light of archival documents. *The Translator* 23:1, 31–48.

Pavlović, Tanja. 2013. The role of collaborative translation protocols (CTPs) in translation studies. *Jezikoslovlje* 14, 549–563. [verkkoaineisto]. [viitattu 15.12.2017]. Saatavissa: http://hrcak.srce.hr/file/165567.

Pokorn, Nike K. 2005. *Challenging the Traditional Axioms: Translation into a Non-mother Tongue*. Amsterdam: John Benjamins.
Probirskaja, Svetlana. 2009. *Rajankäyntiä. Suomen ja Venäjän kahdenväliset valtiosopimukset käännöstieteellisen avainsana-analyysin valossa*. [verkkoaineisto]. [viitattu 23.12.2017]. Tampere: Tampereen yliopisto. Saatavissa: http://urn.fi/urn:isbn:978-951-44-7629-7.
Probirskaja, Svetlana. 2017. ”Does anybody here speak Finnish?” Linguistic first aid and emerging translational spaces on the Finnish-Russian Allegro train. *Translation Studies* 10:3, 231–246.
Prunč, Erich. 2007. *Entwicklungslinien der Translationswissenschaft: Von den Asymmetrien der Sprachen zu den Asymmetrien der Macht*. Berlin: Frank & Timme.
Pym, Anthony. 2010. *Exploring Translation Theories*. New York (NY): Routledge.
Pöchhacker, Franz. 2004. *Introducing Interpreting Studies*. London: Routledge.
Pöchhacker, Franz (toim.). 2015. *Routledge Encyclopedia of Interpreting Studies*. New York (NY): Routledge.
Pöchhacker, Franz & Kadric, Mira. 1999. The hospital cleaner as healthcare interpreter. A case study. *The Translator* 5:2, 161–178.
Reiß, Katharina & Vermeer, Hans J. 1984. *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: Niemeyer.
Risku, Hanna. 2012. Cognitive approaches to translation. Julkaisussa: Chapelle, Carol A. (toim.) *The Encyclopedia of Applied Linguistics*. [verkkoaineisto]. [viitattu 29.12.2017]. DOI: 10.1002/9781405198431.wbeal0145.
Risku, Hanna. 2014. Translation process research as interaction research: From mental to socio-cognitive processes. *MonTI: Monografías de traducción e interpretación* 1, 331–353. [verkkoaineisto]. [viitattu 22.10.2017]. Saatavissa: http://www.e-revistes.uji.es/index.php/monti/article/view/1711.
Risku, Hanna, Rogl, Regina & Milosevic, Jelena. 2017. Translation practice in the field. Current research on socio-cognitive processes. *Translation Spaces* 6:1, 3–26.
Rouhe, Olga. 2017. *Kieliopin rakenne-erojen vaikutus suomi–venäjä-simultaanitulkkauksen sujuvuuteen*. Publications of the University of Eastern Finland. Dissertations in Education, Humanities, and Theology 102. Joensuu: Itä-Suomen yliopisto. [verkkoaineisto]. [viitattu 23.12.2017]. Saatavissa: http://urn.fi/URN:ISBN:978-952-61-2434-6.
Ruokonen, Minna. 2016. Realistic but not pessimistic: Finnish translation students' perceptions of translator status. *The Journal of Specialised Translation* 25, 188–212. [verkkoaineisto]. [viitattu 18.12.2017]. Saatavissa: http://www.jostrans.org/issue25/art_ruokonen.pdf.
Saldanha, Gabriela & O'Brien, Sharon. 2013. *Research Methodologies in Translation Studies.* London: Routledge.
Sandrelli, Annalisa. 2012. Introducing FOOTIE (Football in Europe): Simultaneous interpreting in football press conferences. Julkaisussa: Straniero Sergio, Francesco & Falbo, Caterina (toim.) *Breaking Ground in Corpus-based Interpreting Studies*. Frankfurt: Peter Lang, 119–153.
Seleskovitch, Danica & Lederer, Marianne. 1989. *Pédagogie raisonnée de l'interprétation*. Bruxelles: Didier.
Shlesinger, Miriam. 1998. Corpus-based interpreting studies as an offshoot of corpus-based

translation studies. *Meta* XLIII: 4, 1–8. [verkkoaineisto]. [viitattu 15.12.2017]. Saatavissa: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.119.9115&rep=rep1&type=pdf.

Shreve, Gregory M. & Angelone, Erik (toim.). 2010. *Translation and Cognition*. Amsterdam: John Benjamins.

Szarkowska, Agnieszka, Krejtz, Izabela, Klyszejko, Zuzanna & Wieczorek, Anna. 2011. Verbatim, standard, or edited? Reading patterns of different captioning styles among deaf, hard of hearing, and hearing viewers. *American Annals of the Deaf* 156:4, 363–378.

Taivalkoski-Shilov, Kristiina & Koponen, Maarit. 2017. Contextual factors when reading a translated academic text. Julkaisussa: Alvstad, Cecilia, Greenall, Annjo K., Jansen, Hanne & Taivalkoski-Shilov, Kristiina (toim.) *Textual and Contextual Voices of Translation*. Amsterdam: John Benjamins, 81–99.

Tiittula, Liisa. 2016. Vastaanottajien näkemyksiä ohjelmatekstityksen toimivuudesta. Julkaisussa: Hartama-Heinonen, Ritva, Kivilehto, Marja & Ruokonen, Minna (toim.) *MikaEL Kääntämisen ja tulkkauksen tutkimuksen symposiumin verkkojulkaisu* 9, 87–101. [verkkoaineisto]. [viitattu 18.12.2017]. Saatavissa: https://www.sktl.fi/@Bin/1591116/MikaEL9_Tiittula.pdf.

Tiittula, Liisa & Nuolijärvi, Pirkko. 2013. *Puheen illuusio suomenkielisessä kaunokirjallisuudessa*. Suomalaisen Kirjallisuuden Seuran Toimituksia 1401. Helsinki: Suomalaisen Kirjallisuuden Seura.

Tiittula, Liisa & Nuolijärvi, Pirkko. 2016. Changing norms in translated Finnish fiction: A study of non-standard varieties. *International Journal of Literary Linguistics* 5:3, 1–26.

Tommola, Jorma, Laine, Matti, Sunnari, Marianna & Rinne, Juha O. 2000. Images of shadowing and interpreting. *Interpreting* 5:2, 147–168.

Toury, Gideon. 1995. *Descriptive Translation Studies – and Beyond*. Amsterdam: John Benjamins.

Tuominen, Tiina. 2012. *The Art of Accidental Reading and Incidental Listening: An Empirical Study on the Viewing of Subtitled Films*. Väitöskirja. Tampere: Tampere University Press. [verkkoaineisto]. [viitattu 23.12.2017]. Saatavissa: http://urn.fi/URN:ISBN:978-951-44-9008-8.

Vik-Tuovinen, Gun-Viol. 2006. *Tolkning på olika nivåer av professionalitet*. Acta Wasaensia 153. Väitöskirja. Vaasa: Vaasan yliopisto.

Vuorikoski, Anna-Riitta. 1996. Audience Response to Simultaneous Interpreting. Lisensiaatintyö. Käännöstiede (englanti), Tampereen yliopisto.

Vuorikoski, Anna-Riitta. 2004. *A Voice of its Citizens or a Modern Tower of Babel? The Quality of Interpreting as a Function of Political Rhetoric in the European Parliament*. Acta Universitatis Tamperensis 985. Tampere: Tampereen yliopisto. [verkkoaineisto]. [viitattu 23.12.2017]. Saatavissa: http://tampub.uta.fi/bitstream/handle/10024/67348/951-44-5878-8.pdf?sequence=1&isAllowed=y.

Wadensjö, Cecilia. 1998. *Interpreting as Interaction*. London: Longman.

Walker, Larry. 2017. Communality and conflict: an examination of 'voice' in the Knopf translation of Mishima Yukio's *Spring Snow*. Julkaisussa: Taivalkoski-Shilov, Kristiina, Tiittula, Liisa & Koponen, Maarit (toim.) *Communities in Translation and Interpreting*. Vita traductiva 9. Montréal: Éditions québécoises de l'œuvre, 159–182.

# Luku 24 Psyko- ja neurolingvistiikka

*Leena Maria Heikkola*
https://orcid.org/0000-0002-1795-7656

*Seppo Vainio*

**Mikä?**

Psykolingvistiikka on tutkimusalue, jossa tutkitaan ihmisen kielen ja mielen välistä suhdetta. Neurolingvistiikassa puolestaan tarkastellaan ihmisen aivotoimintojen ja kielen välistä suhdetta. Sekä psyko- että neurolingvistiikka ovat kokeellisia tieteenaloja. Erityisesti näillä aloilla tutkitaan kokeellisin menetelmin puhutun ja kirjoitetun kielen ymmärtämistä, tuottamista ja oppimista.

**Katso myös:**

Luku 1 Kielitaju kielentutkijan työkaluna

Luku 4 Kokeellinen fonetiikka

Luku 17 Kieliopin tutkimus ja kielioppiteoriat

# 1. Johdanto

Psykolingvistiikka ja neurolingvistiikka ovat kokeellista kielentutkimusta (ks. myös Kokeellinen fonetiikka tässä kirjassa). Kokeellisen kielentutkimuksen tutkimuskohteena on kielen prosessointi, jolla viitataan siihen, miten ihmiset ymmärtävät kielellisiä viestejä ja tuottavat kieltä ilmaistakseen ajatuksiaan ja tunteitaan. Alalla tutkitaan esimerkiksi, miten kieltä ymmärretään ja tuotetaan tai miten kielen prosessointi kehittyy lapsilla tai vieraan kielen oppijoilla.

Psykolingvistiikassa tarkastellaan kielen ja mielen suhdetta eli sitä, miten ihmiset omaksuvat, käyttävät, ymmärtävät ja tuottavat kieltä. Esimerkiksi voidaan tutkia, miten sanat on säilötty mielen sanastoon eli mentaalileksikkoon, jolla tarkoitetaan ihmisen mielessä olevaa sanojen ja kieliopillisten rakenteiden varastoa. Kaksikielisillä voidaan tutkia, muodostavatko molempien kielten sanat mielessä yhden suuren leksikon vai kaksi erillistä leksikkoa. Psykolingvistiikassa ollaankin kiinnostuneita siitä, miten kieltä prosessoidaan ja miten kieli järjestyy mielessä. Neurolingvistiikassa puolestaan tarkastellaan kielen ja aivojen suhdetta. Siinä tutkitaan aivomekanismeja, jotka osallistuvat kielen oppimiseen, tuottamiseen, ymmärtämiseen ja käyttöön. Esimerkiksi voidaan tarkastella, mitkä aivojen osat aktivoituvat, kun henkilö kuuntelee ja ymmärtää äidinkielensä puhuttuja lauseita. Neurolingvistiikassa kielen prosessointia tarkastellaan usein erilaisten kielihäiriöiden näkökulmasta (Obler 1999).

Sekä psyko- että neurolingvistiikassa tutkitaan kielen prosessointia erilaisin kokeellisin menetelmin, ja tutkimuskohde voi vaihdella pienistä yksityiskohdista suuriin kokonaisuuksiin, foneemista tekstiin ja diskurssiin. Tutkimusta ohjaavat erilaiset kielen prosessoinnin teoriat, jotka puolestaan perustuvat kielitieteellisiin ja psykologisiin teorioihin. Kokeellisessa kielentutkimuksessa tarkastellaan kielen prosessointia pääasiassa äidinkielen puheen ja kirjoitetun tekstin ymmärtämisen, kielen tuottamisen, kaksikielisyyden sekä vieraan kielen oppimisen näkökulmista. Usein jotakin teemaa tarkastellaan vertaillen eri kieliä, jotta saadaan selville, onko kyseessä yleiseen kielen prosessointiin liittyvä asia

vai onko kyse kielikohtaisesta erityispiirteestä. Kokeellisen kielentutkimuksen koehenkilöt (tästä eteenpäin KH) voivat olla aikuisia, lapsia, vieraan kielen oppijoita, kaksi- tai monikielisiä, jonkin erikoisryhmän jäseniä tai ihmisiä, joilla on jokin sairaus. Valtaosa psykolingvistisestä tutkimuksesta kohdistuu ns. normaaliin kielen prosessointiin, kun taas neurolingvistiikassa tarkastellaan usein neurologisia populaatioita, esimerkiksi MS-tautia tai afasiaa sairastavia. Kokeellinen kielentutkija nojautuu moniin lähitieteisiin, joista tärkeimpiä ovat kielitiede, psykologia, neurotiede, kognitiotiede, neurobiologia ja logopedia. Määrällinen aineistotutkimus on menetelmiltään usein samankaltaista kuin kokeellinen kielentutkimus (ks. tarkemmin luku 3.7.).

Seuraavaksi luvussa 2 esittelemme lyhyesti psyko- ja neurolingvistiikan taustaa ja kielikäsitystä. Luvussa 3 esittelemme psyko- ja neurolingvistiikassa käytettyjä menetelmiä, yleisiä tutkimuskohteita, menetelmien rajoitteita sekä alan lähitieteitä. Lisäksi esitellään tarkemmin psykolingvistiikan ja neurolingvistiikan tutkimusta. Päätämme luvun lyhyeen yhteenvetoon. Luvun lopussa on lisälukemisto, jonka pohjalta lukija voi etsiä lisätietoa.

**Psyko- ja neurolingvistiikan keskeisiä käsitteitä**

**Kielen prosessointi** tarkoittaa tapaa, jolla ihmiset käyttävät kieltä ilmaistakseen kielellisesti ajatuksiaan ja tunteitaan, sekä sitä, miten kommunikaatiota ymmärretään.

**Koehenkilö** (KH) on yksilö, joka osallistuu kokeelliseen tutkimukseen. Jotta koehenkilö voidaan valita tutkimukseen, hänelle tulee kertoa tutkimuksesta ja hänen tulee antaa kirjallinen suostumus osallistumiseen.

**Ärsykkeiden vertaistaminen** tarkoittaa sitä, että käytetyn testin ärsykkeet, esimerkiksi sanat, valitaan niin, että ne vastaavat toisiaan mahdollisimman tarkasti erilaisten tekijöiden suhteen. Tällaisia tekijöitä voivat olla esimerkiksi sanan pituus, sanan yleisyys ja sanan oppimisikä.

**Koehenkilön sisäinen tarkastelu** tarkoittaa, että esimerkiksi morfologisen monimutkaisuuden vaikutusta reaktioaikaan tutkitaan siten, että

koehenkilö saa sekä morfologisesti yksinkertaiset että morfologisesti monimutkaiset sanat tunnistettavakseen, jolloin koehenkilö on itsensä verrokki. **Koehenkilöiden välinen tarkastelu** tarkoittaa, että tutkitaan asiaa, jossa koehenkilö joko kuuluu tai ei kuulu keskeiseen ryhmään. Tyypillisimmillään kyse on interventiomenetelmän toimivuuden testaamisesta. Intervention tehoa tutkitaan siten, että satunnaistetaan koehenkilöt eri ryhmiin, joista vain yksi ryhmä saa intervention, esim. heille opetetaan kieltä tietyllä opetusmenetelmällä. Toinen eli kontrolliryhmä saa esimerkiksi perinteistä kielenopetusta tai ei saa opetusta lainkaan.

**Visuaalinen tehtävä** tehdään näköaistin ja **auditiivinen tehtävä** kuuloaistin perusteella.

**Formalistinen kielikäsitys** perustuu lainalaisuuksien löytämiseen kielestä, **funktionaalinen kielikäsitys** kielen käyttöön ja **kognitiivinen kielikäsitys** siihen, että kieli on yksi kognition osa-alue.

**Verrokki** on koeasetelmassa tiettyyn koehenkilöön tai ryhmätasolla taustamuuttujiltaan vertaistettu osallistuja.

**Hypoteesi** tarkoittaa kokeellisessa tutkimuksessa testattavaa oletusta tutkimusasetelman tuloksesta. Hypoteeseja on vähintään kaksi: nollahypoteesi, jonka mukaan mitattavien muuttujien välillä ei ole yhteyttä, ja tutkimushypoteesi, jossa oletetaan muuttujien välille yhteys, esim. ”Morfologinen monimutkaisuus hidastaa reaktioaikaa”.

**Riippumattomat muuttujat** (tekijä, variaabeli) ovat asetelmassa olevia muuttujia, jotka joko vakioidaan tai joita manipuloidaan tilanteen mukaan. Riippumaton vakioitu muuttuja voi olla esim. sanan esiintymisyleisyys ja manipuloitu muuttuja sanan morfologinen monimutkaisuus. **Riippuvat muuttujat** ovat tutkimuksen selitettäviä asioita, joista halutaan saada selvyyttä, esim. reaktioaika.

**Vastabalansoitu koeasetelma** on tyypillisesti käytössä tilanteissa, joissa koehenkilölle ei voi esittää kaikkia koetilanteita, koska tällöin tulisi liikaa toistoa. Tällaisissa tapauksissa koehenkilöt jaetaan ryhmiin ja he saavat saman määrän ärsykkeitä mutta eri koetilanteista. Jos esim. tutkitaan assosiaation vaikutusta sanantunnistusaikaan, voidaan käyttää

pareja, kuten herätesana ”kissa” ja kohdesana ”koira” tai ”pöytä”. Toinen ärsykeryhmä voisi olla herätesana ”ahma” ja kohdesana ”poro” tai ”kivi”. Ryhmä 1 saisi parit ”kissa” > ”koira” ja ”ahma” > ”kivi”, ryhmä 2 parit ”kissa” > ”pöytä” ja ”ahma” > ”poro”.

**Pilotointi** tarkoittaa koeasetelman teknistä testaamista käytännössä eli pilottikokeiden ajoa. Pilotoinnissa viimeistään nähdään, eteneekö koetilanne suunnitelmien mukaisesti.

**Temporaalinen resoluutio** tarkoittaa sitä, kuinka tarkka aivokuvantamismenetelmä on ajallisesti, ja **spatiaalinen resoluutio** sitä, kuinka tarkka se on sijainniltaan.

# 2. Kokeellisen kielentutkimuksen taustaa

## 2.1. Historiallista taustaa

Psyko- ja neurolingvistiikka ovat saaneet alkunsa 1800-luvulla, mutta tämänhetkisen muotonsa kokeellinen kielentutkimus on saavuttanut 1980–1990-luvuilla. Aivokuvantamisen ja tietokoneiden kehittymisen myötä ala on nopeassa kasvussa.

Wilhelm Wundt (1832–1920) lienee merkittävin yksittäinen kokeellisen kielentutkimuksen kehittäjä, ja häntä luonnehditaankin usein kokeellisen psykologian isäksi (McLeod 2008). Wundt perusti ensimmäisen kokeellisen tutkimuksen ja psykolingvistiikan laboratorion Leipzigin yliopistoon vuonna 1879. Wundt tutki mieltä ja ajattelua systemaattisella tavalla, jossa objektiiviset mittarit ja kokeellisten asetelmien tarkka kontrollointi olivat keskeisessä asemassa (mt.). Voidaankin sanoa, että Wundt käytti kieltä keinona tutkia mieltä. Wundt kirjoitti aiheista, jotka ovat yhä nykyäänkin ajankohtaisia, kuten kielen oppiminen, ymmärtäminen ja tuottaminen, viittomakieli ja lukeminen (Levelt 2014). Wundtin *Die Sprache* vuodelta 1900 (Wundt 1975) oli ainutkertainen kielitieteet yhdistävä teos, joka lopetti vuosisataisen väittelyn

siitä, miten kieli syntyy mielessä (mt.). Vaikka termin *psykolingvistiikka* (engl. *psycholinguistics*) otti ensimmäisenä käyttöön Jacob Kantor vuonna 1936 (mt.), psykolingvistinen tutkimus lisääntyi voimakkaasti vasta 1960-luvulla, erityisesti tietokoneiden käyttöönoton jälkeen.

Neurolingvistiikka on puolestaan kehittynyt 1800-luvulla alkaneesta afasian tutkimuksesta. *Afasialla* viitataan aivoperäisiin kielihäiriöihin (ks. tarkemmin Klippi, Korpijaakko-Huuhka & Lehtihalmes 2017). Afasian tutkimuksessa (engl. *aphasiology*) yritetään selvittää, mitkä aivomekanismit osallistuvat esimerkiksi kielen tuottamiseen ja ymmärtämiseen tarkastelemalla aivovaurioiden vaikutusta kielen prosessointiin. Ranskalainen kirurgi Paul Broca (1824–1880) toi ensimmäisten joukossa esiin yhteyden tietyn aivoalueen ja kielen prosessoinnin välillä. Broca teki ruumiinavauksia henkilöille, joilla oli ollut puhehäiriöitä, ja hän huomasi, että useilla heistä oli aivovaurio vasemmassa etulohkossa. Aivoalue tunnetaan nykyään Brocan alueena. Lääkäri Carl Wernicke (1848–1905) ehdotti myöhemmin, että aivojen eri osa-alueet ovat erikoistuneet erilaisiin kielellisiin toimintoihin: Brocan alue vastaa puheen motorisesta tuottamisesta ja Wernicken aivoalueena nykyään tunnettu alue auditiivisesta puheen ymmärtämisestä. Afasian tutkimus ja ajatus siitä, että kieltä voi tutkia aivojen fyysisten piirteiden kautta, vakiintuivat Brocan ja Wernicken löytöjen pohjalta. 1900-luvun alussa neurologi Korbinian Brodmann (1868–1918) kartoitti aivot niiden pintarakenteen perusteella. Brodmannin alueita käytetään edelleen laajasti neurotieteissä. *Neurolingvistiikka*-termin toi käyttöön Alexandr Luria 1940- ja 1950-luvuilla teoksellaan *Problems in Neurolinguistics*. Harry Whitaker popularisoi neurolingvistiikan Yhdysvalloissa 1970-luvulla perustamalla journaalin *Brain and Language* vuonna 1974.

## 2.2. Psyko- ja neurolingvistiikan kielikäsityksestä

Psyko- ja neurolingvistiikan tutkijat nojautuvat erilaisiin ja keskenään ristiriitaisiinkin kielikäsityksiin. Yleisimmät kielikäsitykset ovat kuitenkin lähtökohdiltaan formalistisia, funktionaalisia, kognitiivisia tai näiden yhdistelmiä. Formalismissa pyritään löytämään kielen lainalaisuuksia,

funktionalismissa tarkastellaan kieltä sen käytön lähtökohdista, ja kognitivismissa kieltä tarkastellaan puolestaan yhtenä kognition osana. Merkittävin jakolinja eri suuntausten välillä liittyy niin sanotun generatiivisen kielitieteen formalistisiin näkemyksiin, jotka eivät ole yleisesti hyväksyttyjä, ja toisaalta funktionaalisen kielitieteen ja kognitiivisen kielitieteen välisiin näkemyseroihin (ks. Kieliopin tutkimus ja kielioppiteoriat tk.). Kokeellisen kielentutkimuksen kielikäsitystä ei useinkaan eksplisiittisesti ilmaista tutkimusraporteissa, joten lukijan tulee olla tarkkana ja sijoittaa tutkimus itse tietynlaisen kielikäsityksen piiriin.

# 3. Psyko- ja neurolingvistiikan tutkimus

## 3.1. Kokeellisen kielentutkimuksen perusteita

Kokeellinen kielentutkimus on ensisijaisesti hypoteesin testausta. Tämä tarkoittaa sitä, että tutkijalla pitää olla selvät tutkimuskysymykset, joihin tutkimuksella pyritään saamaan vastauksia. Tutkimusasetelma suunnitellaan siten, että kahden tai useamman KH-ryhmän tai koetilanteen välille oletetaan tilastollisesti merkitsevä ero (tilastollisesta tutkimuksesta tarkemmin ks. Nummenmaa 2009). Jos tutkimuksen hypoteesissa oletettu ero ei ole tilastollisesti merkitsevä, kyseessä on nollatulos. Tällaisia tuloksia on erittäin vaikea julkaista. Tämä johtuu siitä, että nollatulokseen voi olla myös muita syitä kuin se, että ilmiötä ei ole, esimerkiksi se, ettei koeasetelma ole ollut riittävän tarkka ilmiön esilletuomiseen, tai se, että ilmiö tulee esille jotenkin muuten kuin tutkittavalla tavalla. Hypoteesin testaukseen perustuvaan tarkastelutapaan on kaksi poikkeusta: kliiniset tapaustutkimukset ja kliiniset koeasetelmat. Vaikka psyko- ja neurolingvistisessä tutkimuksessa tarkastellaan kielen prosessointia tai esiintymiä määrällisesti (kvantitatiivisesti), on kielentutkimuksessa aina mukana myös laadullinen (kvalitatiivinen) puoli.

**Kokeellisen kielentutkimuksen koeasetelmatyypeistä ja muuttujista**. Koska kokeellinen kielentutkimus on hypoteesin testausta, tarvitaan hypoteesin testaukseen sopiva asetelma. Hampton (1998) määrittelee

kolme keskeistä ohjetta sille, että kokeellinen asetelma on systemaattinen ja luotettava: manipuloi koeasetelma(t) tehokkaasti, pidä muut merkittävät tilanteet vakioina ja mittaa vaikutus tehokkaasti. Tehokkaan manipuloinnin mahdollistavat hyvin valitut riippuvat ja riippumattomat muuttujat (engl. *dependent/independent variables*). Riippumaton muuttuja on se, jota kokeessa manipuloidaan (esim. sanapituus, frekvenssi, morfologinen rakenne, KH:n ikä tai äidinkieli). Vastaavasti riippuva muuttuja on variaabeli, jonka arvo on riippuvainen riippumattoman muuttujan manipuloinnista. Tutkimuksessa on myös hyvä varmistaa riippuvan muuttujan herkkyys: esimerkiksi reaktioaika ilmaisee virhemäärää herkemmin, kuinka vaikeaa prosessointi on sanantunnistustehtävässä. Tyypillinen koetilanne on sellainen, että yksi tai muutama riippumaton muuttuja muutetaan halutulla tavalla ja muut riippumattomat muuttujat pidetään vakioina. Näin selvitetään se, miten riippumattoman muuttujan (tai riippumattomien muuttujien) manipulointi vaikuttaa yhteen tai useampaan riippuvaan muuttujaan (ks. luku 3.5.).

Esimerkki selventää asiaa. Oletetaan, että tutkija on kiinnostunut siitä, vaikuttaako sanan pituus kirjaimina sen tunnistamisaikaan. Tutkija valmistelee kaksi ryhmää sanoja, jotka hän on vertaistanut esim. sanafrekvenssin ja morfologisen monimutkaisuuden osalta. Vertaistaminen tarkoittaa, että ryhmän sanat ovat valituilta ominaisuuksiltaan samanlaiset eivätkä eroa toisistaan tilastollisesti merkitsevästi. Sanat kuitenkin vaihtelevat sanapituuden suhteen, eli ne ovat joko lyhyitä (esim. *vaja*) tai selvästi pidempiä (esim. *hevonen*). Tarkasteltaessa näiden kahden sanaryhmän tunnistamista keskimääräisen reaktioajan pituuden perusteella voidaan todeta, että sanapituus pidentää sanan tunnistusaikaa. Kokeellisessa asetelmassa tarvitaan siis ainakin kaksi tilannetta: perustilanne (engl. *control condition*) ja koetilanne (engl. *experimental condition*). Koetilanteita voi olla enemmänkin kuin yksi. Kokeeseen sisällytetään yleensä kuitenkin vain vähän koetilanteita, sillä mitä vähemmän koetilanteita on, sitä enemmän koeasetelmassa on tilastollista voimaa.

Kokeenjohtaja pyrkii kokeessaan pitämään kaikki muut kontrolloitavissa olevat tekijät vakioina, jotta vain riippumattoman muuttujan manipulointi erottaa perustilanteen ja koetilanteen. Jotkin kontrolloitavat tekijät liittyvät KH:ihin, kuten kokeessa annettavat ohjeet ja kokeen

samanlainen eteneminen kaikilla. Myös koemateriaali täytyy usein vertaistaa: edellä esitetyssä esimerkissä vertaistettiin esimerkiksi kokeessa käytettyjen sanojen sanafrekvenssi.

Vaikka kokeellinen tutkimus kontrolloitaisiin ja vertaistettaisiin huolellisesti, tulee tulosten tulkinnassa kuitenkin ottaa huomioon se, että laboratorio-olosuhteet voivat vaikuttaa tuloksiin. Tutkimuksen tulosten ekologinen validiteetti eli se, kuinka hyvin menetelmä pystyy antamaan tietoa normaalista kielen prosessoinnista, täytyy arvioida aina tutkimuskohtaisesti. Kokeessa tarkasteltavan vaikutuksen tehokas mittaaminen perustuu siihen, että riippuvat muuttujat ovat riittävän tarkkoja mittaamaan hypoteesien väitteitä, mutta lisäksi riippuvien muuttujien tulee olla riittävän yksinkertaisia. Tällaisia yksinkertaisia muuttujia ovat kokeellisessa kielentutkimuksessa mm. reaktioaika, lukemisaika ja aivosähkökäyrä.

Tutkimusasetelmissa on kaksi perustyyppiä: KH:n sisäinen asetelma (engl. *within-participant design*, Hellier 1998) ja KH:iden välinen asetelma (engl. *between-participant design*, Hampton 1998). KH:n sisäinen asetelma tarkoittaa, että KH on itse itsensä verrokki eli kukin KH saa sekä perustilanteen että koetilanteen. KH:iden välisessä asetelmassa KH-ryhmät ovat eri tilanteissa, joissa toiset saavat koetilanteen ja toiset kontrollitilanteen. Tällaisissa koeasetelmissa järjestetään vastabalansoinnilla yleensä asia niin, että kukin KH saa sekä koe- että kontrollitilanteen ärsykkeitä mutta nämä vaihtelevat esim. kahdessa ryhmässä.

Keinoja maksimoida mahdollisuus saada haluttu efekti koeasetelmalla:

**Moukarimenetelmä** (engl. *sledgehammer approach*). Moukarimenetelmällä tarkoitetaan sitä, että valitaan riippumaton muuttuja, jonka oletettu vaikutus riippumattomaan muuttujaan on mahdollisimman suuri. Tämä takaa vahvan koeasetelman. Esimerkiksi edellä esitetyssä koeasetelmassa valittiin huomattavasti pidempiä sanoja koetilanteeseen kuin perustilanteeseen, jotta sanojen tunnistamiseen kuluva aika (mitattuna reaktioaikoina) olisi mahdollisimman erilainen tilanteiden välillä. Tämä ratkaisu lisää mahdollisuuksia löytää tilastollisia eroja tarkasteltujen tilanteiden välillä.

**Otoskoko**. Mitä suurempi KH-määrä tutkimuksessa on, sitä suurempi mahdollisuus on saada tilastollisesti merkitsevä efekti. Kokeelliset menetelmät ovat usein työläitä, minkä vuoksi tutkimuksen KH-määrät voivat kuitenkin joskus olla pienehköjä. Tähän voi vaikuttaa esimerkiksi se, että KH-ryhmä(t) täytyy vertaistaa sisäisen variaation kontrolloimiseksi; esimerkiksi KH:iden koulutustaustan pitää olla samanlainen.

**Ryhmän sisäisen variaation kontrollointi**. Haasteena voi olla otoksen edustavuus eli se, kuinka hyvin valittu ryhmä edustaa tiettyä yhteiskunnan yhteisöä. Esimerkiksi yliopisto-opiskelijat ovat hyvä ryhmä samankaltaisuuden kannalta mutta haasteellinen edustavuuden näkökulmasta, sillä voi olla, että opiskelijat suoriutuvat tehtävästä paremmin kuin ihmiset keskimäärin. Ryhmän sisäistä variaatiota voi siis kontrolloida valikoimalla ryhmä jostakin tietystä yhteiskunnan yhteisöstä. Ryhmän valinta vaikuttaa tulosten yleistämiseen, joten ryhmä olisi hyvä valita sen mukaan, mikä tutkimuksen tavoite on.

**Riippuvan muuttujan valinta**. Riippuvan muuttujan herkkyys ilmaista haluttua efektiä tulee varmistaa: esim. reaktioaika on virhemäärää herkempi ilmaisemaan prosessoinnin vaikeutta sanantunnistustehtävissä.

**Koetilanteiden määrä**. Kokeeseen sisällytetään vain vähän koetilanteita, sillä mitä vähemmän koetilanteita on, sitä enemmän koeasetelmassa on tilastollista voimaa.

Tutkimuksen eettisyys on keskeistä psyko- ja neurolingvistiikan tutkimuksessa (tarkemmin kielentutkimuksen etiikasta ks. Johdannon luku 4 tk.). Tutkijan tulee luonnollisesti myös taata KH:iden anonyymiys sekä tutkimuksen raportoinnissa että aineiston säilytyksessä. Erityisesti neurolingvistiikassa tutkitaan usein potilasryhmiä, joita koskevat tarkemmat tutkimuseettiset säännöt sekä potilaslain asettamat vaatimukset. Tällöin tutkimukseen hankitaan tutkimuslupa sekä yliopiston omalta että alueen sairaanhoitopiirin eettiseltä neuvostolta.

## 3.2. Kokeellisen kielentutkimuksen eteneminen

Kokeellinen tutkimus koostuu tyypillisesti seuraavista vaiheista: tutkimusaiheen valinta ja taustoitus, koeasetelman hahmottelu ja hypoteesien muodostaminen, ärsykevalinta, koeasetelman tekninen kokoaminen ja kokeen pilotointi, kokeiden ajo ja datan siivous, tulosten tilastollinen tarkastelu sekä tulosten julkaiseminen. Käsittelemme seuraavassa kutakin vaihetta erikseen.

**Tutkimusaiheen valinta ja taustoitus**. Kokeellinen tutkimus alkaa tutkimusaiheen valinnalla. Tutkimusaiheesta kiinnostuminen voi tapahtua eri tavoin. Kokeneempi tutkija voi esitellä aihetta ja siihen liittyvää asiaa, konferenssiesitelmä voi antaa ideoita tutkimusaiheeksi, artikkelia lukiessa tulee idea tutkimusaiheesta, toinen tutkija ehdottaa yhteistyötä aiheen tutkimisessa, kollega voi lähettää aihetta käsittelevän tai sitä sivuavan artikkelin sähköpostilla tai aihe voi sisältyä kokeneen tutkijan rahoitettuun tutkimusprojektiin, johon nuori tutkija pääsee mukaan. Kun aihe on alustavasti valittu, pitää selvittää, onko aihetta tai aihepiiriä tutkittu aiemmin. Tässä vaiheessa haetaan internetistä tai aihepiirin keskeisistä artikkeleista tietoa siitä, onko kyseistä aihetta tutkittu, ja jos on, millaisilla menetelmillä. Oikeiden avainsanojen käyttö auttaa löytämään relevantit artikkelit.

**Koeasetelman hahmottelu ja hypoteesien muodostaminen**. Kun tutkimuksen aihe on valittu ja aiheeseen tutustuttu, on aika hahmotella koeasetelmaa: miten aihetta tarkastellaan ja miten koetilanteita voidaan manipuloida. Toisin sanoen pohditaan, millaisia eri tilanteita koeasetelman pitäisi sisältää, jotta aihetta ja sen vaikutuksia kielenprosessointiin voitaisiin tutkia mahdollisimman yksiselitteisesti. Jos aihetta tutkitaan ensimmäistä kertaa, on parasta tehdä mahdollisimman vahva manipulointi eli ns. moukariasetelma. Jos halutaan tutkia esimerkiksi sitä, missä määrin sanapituus vaikuttaa prosessointiin, lyhyiden ja pitkien sanojen eroksi valitaan monta kirjainta, ei vain yhtä tai kahta. Hypoteesi muodostetaan yleensä aiemman tutkimustiedon pohjalta: se on tutkijan ehdottama selitys tutkittavalle ilmiölle. Ilman hypoteesia tutkimus on tyhjän päällä, eivätkä tutkimuksen tulokset tällöin oikeastaan kerro mitään. Kokeellisessa tutkimuksessa hypoteesin tulee olla sellainen, että

sitä voidaan testata kokeellisesti. Hypoteesi voi joko saada vahvistavaa näyttöä tuekseen tai osoittautua vääräksi. Nollahypoteesi on esimerkiksi se, että kahden tarkasteltavan tilanteen välillä ei ole eroa. Hypoteeseista on myös hyvä keskustella kollegan tai kollegoiden kanssa, sillä joskus aihe on niin innostava, että oma ajattelu voi olla vinoutunut, ja on hyvä kuulla ulkopuolisen näkemyksiä aiheen tutkimisesta ja hypoteesien täsmällisyydestä. Myös koemenetelmä valitaan tässä vaiheessa, jos se ei ole selvä esimerkiksi hypoteesin luonteen vuoksi. Menetelmä voi olla esimerkiksi silmänliikemittaus.

**Ärsykevalinta.** Kun hypoteesit on muodostettu ja koeasetelma selvitetty, alkaa kokeellisen kielentutkimuksen tyypillisesti aikaa vievin ja tärkein osuus eli ärsykemateriaalin valinta ja vertaistus. Kokeen aikana ilmestyneisiin vaikeuksiin voidaan reagoida, mutta huonosta ärsykevalinnasta ei pelasta mikään. Tyypillinen aloittelevan kokeentekijän virhe on tarkkuuden puute, joka ilmenee yleensä siten, että aikapaineessa koemateriaalin kerääjä ei malta olla riittävän huolellinen ja perinpohjainen ärsykevalinnassa. Usein syy tähän on, että kokeentekijää avustava kokenut tutkija ei ole riittävästi korostanut huolellisuuden merkitystä ja opastanut ärsykevalinnan koukeroissa. Toinen yleinen virhe on, että kokematon tutkija pyrkii liittämään koeasetelmaan liian monta hypoteesia, jolloin seurauksena on liian pieni määrä ärsykkeitä kutakin hypoteesia kohden. Tällainen tilanteiden vähäisyys heikentää tulosanalyysien tilastollista voimaa, jolloin mahdollisuus halutun efektin todentamiseen vähenee. Nyrkkisääntönä voi pitää sitä, että on parempi tehdä useita kokeita kuin pyrkiä tekemään kaikki mahdollinen yhdellä kertaa. Näin on siksi, että jos ärsykkeitä on riittävästi kokeen moniin hypoteeseihin, niin koe on todennäköisesti pitkä ja raskas sekä kokeenjohtajalle että erityisesti KH:ille. KH:iden väsyminen ei ole toivottavaa, koska väsymys vaikuttaa herkästi suoritukseen.

On tärkeää huomata myös se, että jo ärsykevalintaa tehtäessä tutkijalla pitää olla käsitys siitä, miten tuloksia analysoidaan tilastollisesti. Tarkastelutapa vaikuttaa nimittäin siihen, miten ärsykeaineisto koostetaan. Psykolingvistiikan viitekehyksessä voidaan esittää laadullisia tutkimustuloksia, mutta edustamamme kokeellinen kielentutkimus on pääasiallisesti määrällistä eli kvantitatiivista tutkimusta, jossa ilmiö joko on tai ei ole sen

mukaan, saadaanko tilastollisesti merkitsevä tulos. Toisin sanoen pelkkä ero tilanteiden välillä ei merkitse mitään, jos se ei ole tilastollisesti merkitsevä. Laadullinen tarkastelu voi kuitenkin joskus olla tarpeen, esimerkiksi koeasetelmaa suunnitellessa. Jos esimerkiksi haluaa tutkia partisiippien prosessointia, pitää selvittää sekä se, millaiset sanat määritellään partisiipeiksi, että se, millaisissa tilanteissa partisiippi on leksikaalistunut adjektiiviksi. Näitä ei voida määrällisellä tarkastelulla selvittää.

**Koeasetelman tekninen kokoaminen ja kokeen pilotointi**. Ärsykevalinnan jälkeen on vuorossa koeasetelman tekninen toteutus ja kokeen pilotointi. Tässä vaiheessa koelaitteistoon säädetään oikeat asetukset ja lisätään ärsykeaineisto sekä tarkistetaan koeohjelma. Helpoimmillaan ohjelmointia ei tarvitse muuttaa, vaan laitteiston säätöjen muuttaminen riittää, mutta koeasetelma voi myös edellyttää aivan uudenlaista ohjelmaa. Joskus jotakin kokeen osaa täytyy testata, jotta varmistetaan, että ohjelma toimii halutusti. Kun kaikki on kohdallaan, tehdään testiajot eli koe pilotoidaan. Pilottivaiheessa testataan muutamia KH:itä, joiden tuloksia ei kuitenkaan oteta mukaan varsinaiseen analyysiin. Muutaman testin jälkeen tiedetään, toimiiko ohjelma halutusti, ja jos näin on, tarkastellaan tulostiedostoja, jotta nähdään, onko tarpeellinen data tallentunut. Jos kaikki on hyvin, edetään varsinaisen kokeen ajoon. Tässä vaiheessa rekrytoidaan tarvittavat KH:t kokeeseen.

**Kokeiden ajo ja datan siivous**. Kun koeasetelma on testattu ja todettu toimivaksi, voidaan siirtyä tutkimuksen varsinaiseen testausvaiheeseen eli kokeiden ajoon. Testausvaiheessa testataan etukäteen sovittu määrä KH:itä, minkä jälkeen siivotaan eli poistetaan virheellinen data, kuten häiriöt silmänliikekameran rekisteröinnissä laitteiston toimiessa muuten normaalisti. Tässä yhteydessä – asetelmasta riippuen – voidaan tarvittaessa poistaa myös tilastollisin kriteerein poikkeavat havainnot, esimerkiksi arvot, jotka poikkeavat yli kolme keskihajontaa kyseisen datan keskiarvosta. Tällä siivouksella saadaan data yhtenäiseksi, jolloin poikkeukselliset tulokset eivät vinouta tai peitä tuloksia.

**Tulosten tilastollinen tarkastelu**. Tulokset käsitellään tilastollisesti. Tässä luvussa ei tarkastella tilastollisia menetelmiä, koska niistä on kirjoitettu laajasti muualla (ks. esim. Nummenmaa 2009; Grönroos 2008; myös Määrällinen korpuslingvistiikka tk.). Mainittakoon, että tyypillinen

kokeellisen kielentutkimuksen datan tilastollinen tarkastelutapa on ollut varianssianalyysi (ANOVA), mutta nykyään dataa tarkastellaan enenevässä määrin monitasomallinnuksella (engl. *linear-mixed-modelling*).

**Tulosten julkaiseminen**. Kun varsinainen tutkimus on tehty, on aika raportoida tulokset. Koeasetelmia selostava tieteellinen artikkeli on tekstilajina yleensä määrämuotoinen: siihen kuuluu tyypillisesti johdanto, teoreettisen taustan esittely, metodiosuus, tulokset, pohdinta ja päätelmät. Kirjoittamisprosessiin kuitenkin vaikuttaa olennaisesti, mihin julkaisuun käsikirjoitus lähetetään, koska eri julkaisujen kirjoitusohjeet vaihtelevat esim. viittaustavan, rakenteen ja sanamäärän suhteen. Lähes poikkeuksetta kokeellisen kielentutkimuksen raportointi aloitetaan metodiosuudesta, jossa kuvataan tutkimusaineisto sekä esitellään KH:t ja koeasetelma. Tämän jälkeen kirjoitetaan auki tulokset. Lopuksi voi siirtyä esittelyosuuteen (engl. *Introduction*) ja pohdintaan (engl. *Discussion, General Discussion*). Joissakin julkaisuissa vedetään yhteen tulokset ja niiden implikaatiot vielä lyhyesti päätelmät-osiossa (engl. *Conclusions*). Tieteellisen artikkelin kaksi tärkeää tekijää ovat lisäksi abstrakti ja avainsanat (engl. *keywords*). Ne vaikuttavat mitä suurimmassa määrin siihen, kuinka monet lukevat artikkelin ja kuinka paljon siihen viitataan muissa tutkimuksissa.

## 3.3. Kokeellisen kielentutkimuksen menetelmiä

Kokeellisen tutkimuksen menetelmät voidaan jaotella esimerkiksi modaliteetin, suoritustavan tai tarkasteltavan populaation mukaan. *Modaliteetilla* viitataan siihen, minkä aistin varassa tutkimuksen tehtävä tehdään. *Suoritustavan* mukaista jaottelua kuvataan englanninkielisessä kirjallisuudessa jakamalla menetelmät online- ja offline-menetelmiin: *online-menetelmillä* viitataan kielenprosessoinnin tarkasteluun reaaliajassa, *offline-menetelmillä* tarkastellaan kielenprosessointia ei-ajallisesti, esimerkiksi tarkkuuden tai virheiden näkökulmasta. Populaatioista kokeellisessa tutkimuksessa tarkastellaan pääasiassa ryhmiä, mutta 2000-luvulla myös kokeelliset yksittäistapaustutkimukset (engl. *single study experimental design*, SSED) ovat yleistyneet.

### 3.3.1. RYHMÄ- JA TAPAUSTUTKIMUS

**Kokeellinen ryhmätutkimus**. Kokeellisessa ryhmätutkimuksessa tarkastellaan tietyn ryhmän kielellistä ja usein myös kognitiivista suoriutumista suhteessa terveistä verrokkihenkilöistä koostuvaan ryhmään. Verrokkiryhmän tulee vastata KH-ryhmää taustamuuttujiltaan (mm. ikä, sukupuoli, koulutus) tarkasti. Neurolingvistiikassa tarkastellaan pääasiassa patologisia ryhmiä, esim. henkilöitä, joilla on afasia tai jokin kehityksellinen kielihäiriö. Kokeellisen kielentutkimuksen piirissä tutkitaan myös paljon toisen kielen oppijoita, esimerkiksi sitä, miten L2-puhujien kielitaito eroaa L1-puhujista. Ryhmätutkimuksissa molemmat ryhmät osallistuvat tutkimuskysymyksen perusteella valittuihin kielellisiin ja mahdollisesti kognitiivisiin testeihin. Näiden testien vastauksia verrataan ryhmätasolla, eli ryhmätutkimuksissa ei yleensä kiinnitetä huomiota ryhmän sisällä esiintyvään yksilölliseen vaihteluun. Ryhmätasoista tarkastelua on myös kritisoitu, ja nykyään myös ryhmien sisäistä vaihtelua raportoidaan enemmän tutkimuksissa. Koska tuloksia kuitenkin tarkastellaan pääasiassa ryhmätasolla, täytyy tutkimuksen ryhmien olla suurehkoja (15–100 KH:ä per ryhmä), koska ryhmiä vertaillaan useimmiten tilastollisesti.

**Kokeellinen tapaustutkimus**. Tapaustutkimuksessa tarkastellaan yhden tai muutaman KH:n tiettyä kielellistä käyttäytymistä syvemmin kuin ryhmätason tutkimuksissa (ks. tarkemmin Beeson & Robey 2006; Byiers, Reichle & Symons 2012). Tarkastelu on usein laadullisempaa, koska KH:itä on vain muutama. Tämä lähestymistapa sopii erityisen hyvin tapauksiin, joita esiintyy hyvin vähän (esimerkiksi monikieliset afaatikot). Usein tapaustutkimukset ovat kuvailevia, mutta tapaustutkimus voi olla myös kokeellinen, jos tarkastellaan esimerkiksi jonkin tietyn intervention vaikutusta KH:n kieleen. Joskus tapaustutkimuksissa vertaillaan yhden tai kahden KH:n kielellisiä tai kognitiivisia tuloksia vertaistettuun verrokkiryhmään. Verrokkiryhmä on tarpeellinen erityisesti silloin, kun käytettyjen testien tuloksista tarkastellaan reaktioaikoja tai muita aikaan sidottuja arvoja.

### 3.3.2. OFFLINE-MENETELMIÄ

Offline-menetelmillä tarkastellaan kielenprosessointia ei-ajallisesti. Niiden avulla voidaan selvittää esimerkiksi KH:iden kielen tuottamista, arvioita esimerkiksi lauseen luonnollisuudesta tai kieliopillisuudesta tai kielellistä muistia. Keskeistä offline-menetelmissä on, että niissä ei tarkastella KH:iden tuotoksen kestoa. Ajankäyttö voi kuitenkin olla tehtävässä rajattu. Online-kokeissa käytettävää materiaalia pilotoidaan usein teettämällä offline-testejä, joissa tarkastellaan esimerkiksi kokeessa käytettävien lauseiden semanttisia ominaisuuksia ja lauseiden luontevuutta. Pyrkimyksenä on saada mahdollisimman hyvin vertaistettuja aineistoja kokeellista tutkimusta varten tai saada tarkkaa tietoa esim. siitä, miten hyvin luettu teksti muistetaan lukemisen jälkeen. Offline-menetelmien tehtävätyyppejä esitellään taulukossa 1.

| | |
|---|---|
| **Yksittäisten sanojen normitus** | Tässä tehtävätyypissä KH:t arvioivat tyypillisesti sanojen ominaisuuksia, kuten tuttuutta, konkreettisuutta, kuviteltavuutta, edustavuutta ja tyypillisyyttä/prototyyppisyyttä. |
| **Assosiaatiotehtävät** | Assosiaatiotehtävissä KH:t arvioivat, kuinka hyvin jotkin sanat tai lauseenosat sopivat yhteen. |
| **Tuottotehtävät** | Tuottotehtävissä KH:t joko puhuvat tai kirjoittavat. Tuottoa voidaan ohjailla (esim. ”tuota sopiva substantiivi”) tai siinä voidaan käyttää vapaata assosiaatiota (esim. ”tuota viisi ensimmäistä mieleesi tulevaa sanaa”). |
| **Arviointitehtävät** | Arviointitehtävissä KH:t arvioivat annetun esimerkin erilaisia piirteitä, esimerkiksi lauseen luonnollisuutta määrätyllä asteikolla tai sitä, miten hyvin lauseen viimeinen sana sopii edeltävään osuuteen. |
| **Muistamistehtävät** | Muistamistehtäviä on kaksi pääversiota: välitön ja viivästetty. KH:itä pyydetään esimerkiksi tuottamaan tiivistelmä juuri luetusta tarinasta, tai heitä pyydetään tiivistämään tarina viikon kuluttua tarinan lukemisesta. |

Taulukko 1. Offline-menetelmien tehtävätyyppejä.

### 3.3.3. ONLINE-MENETELMIÄ

Online-menetelmillä tarkastellaan KH:n varsinaista suoritusta. Suorituksen (esim. suoritusajan) mittauksen kaksi pääsuuntausta ovat reaktioaikaan liittyvät menetelmät ja jatkuvan mittauksen menetelmät. Reaktioaikamenetelmillä saadaan yksi mittauspiste. Jatkuvan mittauksen menetelmillä voidaan puolestaan tarkastella KH:n toimintaa prosessina.

#### Reaktioaikamenetelmiä

**Visuaalinen sanantunnistus**. Visuaalinen sanantunnistus (engl. *visual word recognition*), toiselta nimeltään leksikaalisen päätöksenteon tehtävä (engl. *visual lexical decision*), on yleisin ja käytetyin reaktioaikamenetelmä. KH:n tehtävä on tunnistaa tietokonekuvaruudulta mahdollisimman nopeasti ja tarkasti, onko kulloinkin kuvaruudulle tuleva sana tutkittavana oleva kielen sana vai ei. Ärsykkeinä käytetään tällöin tietyn kielen sanoja (esim. *tulppa*) ja saman kielen fonotaksin mukaisia epäsanoja (esim. **talppa*). Tyypillisesti tämän tehtävän avulla tutkitaan jotakin kielen yleistä piirrettä (esim. morfologista monimutkaisuutta), jonka vaikutuksia reaktioaikanopeuteen halutaan selvittää. Perusajatuksena on, että mitä kauemmin reaktio kestää, sitä vaativammasta prosessoinnista on kyse. Tämän kuten muidenkin reaktioaikamenetelmien ongelmana on ns. ekologinen validiteetti: voidaanko tuloksia yleistää arkielämään?

**Auditiivinen sanantunnistus**. Auditiivinen sanantunnistus (engl. *auditive word recognition, perceptual identification*) on kuuloon perustuva sanantunnistusmenetelmä, jossa KH kuulee sanoja yksi kerrallaan ja hänen tehtävänsä on tunnistaa kukin sana niin tarkasti kuin mahdollista. Asetelman vaikeusasteen lisäämiseksi ärsykkeiden kuulemista häiritään joskus taustahälyllä, joka voi olla puheen kaltaista tai muunlaista ääntä.

**Nimeäminen**. Nimeäminen (engl. *naming*) tarkoittaa asetelmaa, jossa KH:n pitää mahdollisimman nopeasti toistaa ääneen tietokoneen kuvaruudulla kulloinkin näkyvä sana tai nimetä ruudulla näkyvä kuva. Menetelmällä pyritään saamaan tietoa sanantunnistuksen alkuvaiheen ilmiöistä. Metodilla on kolme vaikeasti ratkaistavaa erityisongelmaa, joita muilla reaktioaikamenetelmillä ei ole. Ensinnäkin toisto on usein nopeampaa kuin ymmärtäminen, joten jos sana on pitkä, henkilö aloittaa

ääntämisen ennen kuin hän on lukenut koko sanan. Tämä on ongelma erityisesti suomessa, jossa sanat ovat usein pitempiä kuin esimerkiksi englannissa. Toiseksi sananalun suhteellinen vaikeus pidentää ääntämisen aloittamisaikaa (ks. Kokeellinen fonetiikka tk.). Kolmanneksi ääntämisen mittauspisteen aloituskohta ei ole yksiselitteinen, vaikka tyypillisesti käytetäänkin ääntämisen alkamishetkeä (määriteltynä esimerkiksi akustisen tarkastelun perusteella). Esimerkiksi ääntäminen voi hidastua loppuosassa. Vaikka kahden vertailtavan sanan ääntämisen alkamiskohta on teoriassa sama, käytännössä KH:t voivat tuottaa sanojen alkuja eri nopeuksilla.

**Heräteasetelma**. Heräteasetelma (engl. *priming*) on asetelma, jossa halutaan saada selville herätteen (engl. *prime*) ja kohteen (engl. *target*) välisen suhteen läheisyys. *Herätteellä* tarkoitetaan ärsykettä, joka vaikuttaa seuraavaan ärsykkeeseen vastaamiseen tiedostamattomasti. Jos tietokoneen kuvaruudulla esitetään esimerkiksi sana *kissa* tai *kana*, *kissa*-sanan jälkeen esitettynä kohdesana *koira* on nopeampi tunnistaa kuin *kana*-sanan jälkeen esitettynä. Ajatuksena on, että kohdesanaan liittyvän herätteen tunnistaminen aktivoi semanttisen kentän mutta kohdesanaan liittymätön tai vain kaukaisesti liittyvä sana ei aiheuta samanlaista kohdesanan tunnistuksen helpottumista (engl. *facilitation*). Heräteasetelmalla voidaan tutkia myös eri aistien välistä herätevaikutusta (engl. *intermodal priming*). Tyypillisimmät asetelmat ovat joko visuaalinen heräte ja visuaalinen kohde tai auditiivinen heräte ja visuaalinen kohde. Kaikille reaktioaikamenetelmille on yhteistä, että niissä voidaan käyttää myös niin sanottua vastaa–älä vastaa -versiota (*go/no-go*), jolla pystytään vaikuttamaan KH:n vastaamistapaan. Vastaa–älä vastaa -versiossa KH:n tehtävänä on reagoida vain ennalta määrätyn tyyppisiin ärsykkeisiin, esim. vain oikeisiin sanoihin, kun tehtävässä esitetään oikeita sanoja ja epäsanoja. Erityisesti heräteasetelmissa on tällä versiolla joskus varsin suuri merkitys.

Yleisesti voidaan todeta, että reaktioaikamenetelmien etuina ovat asetelman ja koesarjan tulosten saannin helppous ja laitteiston helppo saatavuus. Haasteena on tulosten suhde normaaliin kielenprosessointiin: kertooko hitaampi reaktioaika todellakin siitä, että prosessointi on hankalampaa? Sanan yleisyys on kuitenkin merkittävimpiä reaktionopeutta

lisääviä tekijöitä, mikä antaa loogisen perusteen reaktioajan käyttöön prosessoinnin vaativuuden mittarina.

### Jatkuvan mittauksen menetelmiä

Kokeellisessa kielentutkimuksessa käytetään sekä silmänliikemittausta että aivokuvantamisen menetelmiä. Seuraavaksi esitellään näitä jatkuvan mittauksen menetelmiä.

**Silmänliikemittaus.** Silmänliikemittauksessa rekisteröidään KH:n silmänliikkeitä suhteessa ärsykkeeseen. Silmänliikkeiden oletetaan kertovan KH:n tietyn tehtävän monimutkaisuudesta ja sen edellyttämästä prosessoinnista. Silmänliikkeiden rekisteröinti perustuu siihen, että silmät hypähtelevät ja pysähtelevät lukiessa tai mitä tahansa katsellessa. Nämä pysähdysajat ja -kohdat pystytään tietokoneistetulla silmänliikekameralla (engl. *eye tracker*) mittaamaan ja rekisteröimään myöhempää analyysia varten. Silmänliikemittauksessa käytettävä laitteisto koostuu kahdesta tietokoneesta, jotka on yhdistetty toisiinsa ja kahteen pieneen infrapunakameraan. Kuva 1 alla havainnollistaa KH:n näkymän tyypillisessä tekstin prosessoinnin koeasetelmassa.



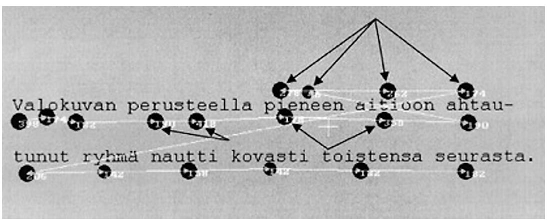
Kuva 1. Silmänliikemittaus (julkaistu SR Research Ltd:n luvalla).

Lukemisen silmänliiketutkimus perustuu kahteen yleiseen periaatteeseen. Ensimmäisen periaatteen mukaan prosessointiin vaikuttaa kohteen kompleksisuus tai muu hankaloittava ominaisuus: prosessoimme jotakin kielen rakennetta sitä kauemmin, mitä vaikeampi se on. Esimerkiksi lauseke luetaan nopeammin sivulauserakenteessa kuin konverbirakenteessa (Äiti ajoi nurmikkoa, ***kun isä hinkkasi*** *auton konepeltiä* vs. Äiti ajoi nurmikkoa ***isän hinkatessa*** *auton konepeltiä*). Toisen periaatteen eli niin sanotun silmä-mielihypoteesin (engl. *eye-mind hypothesis*) mukaan sanojen prosessointi ja lukeminen ovat käytännössä sama asia. Viitatuin ja samalla kritisoiduin versio tästä periaatteesta on Justin ja Carpenterin (1980) niin sanottu välittömyyshypoteesi (engl. *immediacy hypothesis*). Tämän äärimmäisen tulkinnan mukaan sanan prosessointi alkaa heti sen lukemisen alkaessa ja loppuu heti, kun sanan lukeminen loppuu. Se että sanan prosessointi loppuu lukemisen loppumiseen, on kuitenkin osoitettu vääräksi. Rayner ja Duffy (1986) osoittivat ensimmäisinä niin sanotun valumisefektin (engl. *spill over effect*) ja viivästetyn efektin (engl. *delayed effect*). *Valumisefektillä* tarkoitetaan tilannetta, jossa kohdealueen prosessoinnin hidastumisen ohella myös kohdealueen jälkeinen rakenne prosessoidaan hitaammin, eli efekti ikään kuin valuu yli. Viivästetyssä efektissä prosessoinnin vaikeus ei ilmene heti kyseisen rakenteen lukemiseen käytetyssä ajassa vaan vasta sitä seuraavan rakenteen lukemisessa. Välittömyyshypoteesiin nähden toiseen suuntaan äärimmäinen tulkinta prosessoinnista lähtee siitä, että monet kielen tekijät, kuten kongruenssi, eivät vaikuta varsinaiseen lukemiseen vaan niitä käytetään vasta jonkinlaisena tarkistusjärjestelmänä (engl. *back up procedure,* Nicol, Forster & Veres 1997).

Jos luettava lause on hankala prosessoida, lukeminen hidastuu. Tutkimuksessa tarkastellaan tyypillisesti, vaikuttaako tutkittava ilmiö lukemisen nopeuteen. Lukemisen hidastuminen voidaan jakaa erilaisiin alaryhmiin. Näitä ovat tarve hidastaa lukemista vaikeassa kohdassa, tarve hidastaa lukemista hankalan kohdan jälkeisessä rakenteessa tai tarve palata aiempiin kohtiin tekstissä. Luettava kohta voi käsittää mitä tahansa sanan osasta kokonaiseen kappaleeseen.

Yleensä lukemisen vaikeus ilmaistaan alan teknisiä käsitteitä käyttäen, jolloin keskeisenä viitekehyksenä on tapahtumien ajallinen

jatkumo. Tutkittavan alueen ensimmäinen fiksaatio eli ensimmäinen silmäys (engl. *first fixation*) on lyhytaikaisin havaittavissa oleva lukemisen vaativuuden muuttuja. Ensimmäinen silmäys tarkoittaa sitä lukemisaikaa, jonka KH:n ensimmäinen silmäys kestää. Ajallisesti seuraava muuttuja on kohdealueen ensimmäinen lukeminen (engl. *first pass reading, gaze duration*). Tämä on kaikkein keskeisin suure, ja sillä tarkoitetaan aikaa, jonka KH käyttää kohdealueen lukemiseen ennen siirtymistään tekstissä joko eteenpäin seuraaviin sanoihin tai taaksepäin aiempiin sanoihin. Edellä mainitut valumisefekti ja viivästetty efekti ovat mahdollisia eroja kohdealueen jälkeisen rakenteen lukemisessa, ja ne ovat usein aikajanan seuraavia muuttujia. Useimmiten ajallisesti viimeinen muuttuja on kohdealueen uudelleenlukeminen (engl. *regression, rereading*), joka sisältää kaiken sen kohdealueen lukemisen, joka tapahtuu ensimmäisen lukemiskerran jälkeen. Uudelleenlukemisen yhteydessä käytetään mittarina usein myös uudelleenlukemisen todennäköisyyttä (engl. *probability for regression*). Mikäli tätä muuttujaa käytetään, voi muuttuja olla myös kohdealueen kokonaislukemisaika (engl. *total fixation time*), jolla tarkoitetaan kohdealueeseen kohdistuvaa yhteenlaskettua lukemisaikaa. Näiden pääosin ajallisten muuttujien lisäksi myös fiksaation sijaintia voi tarkastella. Tällöin tarkastellaan yleensä toisaalta sanan ensimmäisen fiksaation sijaintia ja sitä edeltävän silmänliikkeen (engl. *saccade*) lähtöpaikkaa, joista saa laskettua sakkadin eli fiksaatiosta toiseen siirtyvän nopean silmänliikkeen pituuden (ks. kuvio 1). Kuviossa 1 uudelleenlukemiset on selvyyden vuoksi esitetty tekstin yläpuolella ja merkitty nuolilla.




Kuvio 1. Silmänliikkeiden rekisteröinnissä käytetyt suureet (Vainio 2006; julkaistu *Puhe ja Kieli* -lehden luvalla).

Menetelmän etuna on, että KH:n tehtävä on luonnollinen: KH vain lukee lauseita tai tekstiä kuvaruudulta samalla, kun hänen silmänliikkeensä rekisteröidään. Toisin sanoen KH ei joudu tekemään mitään keinotekoista, vaan hän lukee kuten muutoinkin lukisi. Tällä perusteella saatuja tuloksia voidaan käyttää argumentoinnissa melko hyvin. Tutkimuksen onnistumiseen kuitenkin vaikuttaa olennaisesti, miten katseen kulloinenkin kohde vastaa prosessoinnin kohdetta ja kuinka hyvin kokeiden ärsykemateriaalit on vertaistettu.

**Funktionaalinen magneettikuvantaminen (fMRI)**. Magneettikuvantaminen (engl. *magnetic resonance imaging*) on pääasiassa aivojen rakenteiden staattinen kuvantamismenetelmä. Magneettikuvantaminen perustuu vesimolekyylien atomiydinten magneettiseen toimintaan. Tämä saadaan aikaan luomalla ulkoinen magneettikenttä, jolloin vesimolekyylien atomit järjestäytyvät linjaan. MRI-kuvat luodaan matemaattisen kaavan avulla kolmiulotteisiksi kuviksi aivoista. Koska MRI-kuvat perustuvat verenkiertoon, niiden temporaalinen resoluutio ei ole kovin tarkka. Tekniikan hyvä puoli on korkea spatiaalinen (eli sijaintiin liittyvä) resoluutio, eli aktivaatiota aivoissa voidaan tarkastella jopa millimetrin tarkkuudella. MRI-tekniikka on ei-invasiivinen, eli KH ei tutkimuksessa altistu mahdollisesti kehoa vahingoittaville toimenpiteille, koska kuvantamistekniikka perustuu magneettikenttiin. Kokeellisessa kielentutkimuksessa käytetään funktionaalista magneettikuvantamista, jolloin aivojen toimintaa voidaan tarkastella kielen prosessoinnin aikana.

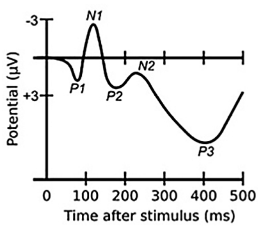
**Magneettienkefalografia (MEG)**. Magneettienkefalografia (engl. *magnetoencephalography*) mittaa aivojen sähköisen toiminnan tuottamia magneettikenttiä. MEG:n spatiaalinen resoluutio on korkea, ja myös sen temporaalinen resoluutio on tarkempi kuin fMRI:llä. MEG on menetelmänä täysin ei-invasiivinen, koska siinä ei käytetä radioaktiivisia aineita, ja se mittaa hermoston toimintaa suoremmin kuin tarkastelemalla aineenvaihduntaa tai verenkiertoa. Magneettikenttiin perustuva MEG sopii erityisen hyvin yksittäisten KH:iden ja yksilöllisen variaation tarkasteluun, koska se ei vaadi ryhmän sisällä tehtävää keskiarvoistamista.

**Elektroenkefalografia (EEG)**. Elektroenkefalografialla (engl. *electroencephalography*) mitataan aivosolujen tuottamaa sähköistä toimintaa. EEG-mittauksessa tarkastellaan kahden pisteen välisiä

sähköpotentiaalien eli aivoaaltojen eroja. Päänahalle asetetaan elektrodeja, jotka mittaavat aivokuoren hermosolujen sähköistä toimintaa eri aivopuoliskoissa ja aivolohkoissa. Mitattu sähköinen toiminta esitetään usein visuaalisesti, ja eri alueiden toimintaa verrataan toisiinsa.

**Herätepotentiaalit (ERP).** *Herätepotentiaaleilla* eli *herätevasteilla* (engl. *event-related potentials*) viitataan sähköiseen toimintaan, joka syntyy, kun keskushermosto reagoi tiettyyn sensoriseen ärsykkeeseen. Nämä potentiaalit ovat suhteellisen pieniä (1–5 mikrovolttia), ja ne rekisteröidään elektrodeilla samoin kuin EEG:tä mitattaessa. Koska potentiaalit ovat hyvin pieniä, ERP-mittauksessa lasketaan useiden saman ärsykkeen aiheuttamien potentiaalien keskiarvoja. Näin saadaan suodatettua pois häiriöitä, kuten taustamelua. ERP on tyypillisesti sarja positiivisia ja negatiivisia huippuja, jotka heijastavat varhaisia tai myöhäisiä informaation prosessoinnin eli tiedonkäsittelyn vaiheita (ks. kuvio 2). Negatiiviset huiput merkitään kirjaimella N ja positiiviset huiput kirjaimella P. Kirjaimeen liitetään numero millisekunneissa, joka kuvastaa aikaa ärsykkeen esittämisestä, esimerkiksi N400 tai P600. Huippuihin voidaan viitata myös järjestysnumeroilla, esim. N1 tai P3, alkaen ärsykkeen esittämishetkestä. Lisäksi on huomattava, että alan artikkeleissa käytetyissä kuvaajissa on usein negatiiviset arvot yläpuolella ja positiiviset arvot alapuolella.

Suhteessa muihin aivojen kuvantamismenetelmiin ERP:n käyttö neurolingvistisessä tutkimuksessa on hyödyllistä, koska menetelmä on edullinen eikä se ole invasiivista. Menetelmän spatiaalinen resoluutio on kohtuullinen ja temporaalinen resoluutio erinomainen, mikä mahdollistaa ajallisesti nopeasti tapahtuvan kognitiivisen ja kielellisen prosessoinnin mittaamisen. (Ahlsén 2006, 161–165.) Neurolingvistisessä tutkimuksessa käytetäänkin usein herätevasteita, joita mitataan EEG:llä.




Kuvio 2. Herätevaste (Commons Wikimedia).

Kielellinen syöte näkyy aaltomaisena kuviona, kun samaa kielellistä ärsykettä koskevista EEG-aalloista lasketaan keskiarvo (ks. kuvio 2). ERP-komponentti on yksi tämän monimutkaisemman, keskiarvoista lasketun ERP-aallon aalloista. Useimmiten ERP-komponenteista tarkastellaan niiden amplitudia eli vahvuutta, latenssia eli viivettä sekä topografiaa eli sijoittumista aivoalueille. ERP-efekti on puolestaan amplitudien erotus kahden eri ERP-komponentin välillä, esimerkiksi kahden eri lauseen välillä.

Seuraavaksi esitellään kielentutkimuksen yleisimpiä ERP-komponentteja ja niiden tutkimuksessa käytettyjä paradigmoja (ks. tarkemmin Swaab, Ledoux, Camblin & Boudewyn 2011). Semanttiseen prosessointiin on kirjallisuudessa liitetty käsitteellistä poikkeamaa heijastava N400 ja syntaktiseen prosessointiin varhaista syntaktista prosessointia heijastava ELAN sekä myöhäistä syntaktista prosessointia heijastava P600 (ks. tarkemmin Swaab, Ledoux, Cambln & Boudewyn 2011).

**N400** on varhaisin löydetty kielen prosessoinnin ERP-komponentti (Kutas & Hillyard 1980). Kyseessä on ERP-komponentti, jota tarkastellaan pääasiallisesti reaktiona semanttiseen tai konseptuaaliseen eli käsitteelliseen poikkeamaan. N400 ilmenee negatiivisena huippuna noin 400 ms kriittisen sanan esittämisen jälkeen. Topografialtaan tämä komponentti on sentroparietaalinen, eli elektrodien amplitudi on vahvin keskisellä päälaella. N400-komponentti heijastaa vaikeutta integroida semanttista informaatiota. Esimerkiksi seuraavanlaisilla lauseilla voidaan tarkastella N400-efektiä:

(1) *Hän levitti leivälle *<u>sukat</u>.*
*Hän levitti leivälle <u>voita</u>.*

Kaikille sisältösanoille syntyy N400, mutta komponentin amplitudi on sitä voimakkaampi, mitä vaikeampi sanaa on integroida edeltävään kontekstiin. Myös epäsanojen N400 on suurempi kuin tavallisten sanojen. N400-komponentissa näkyy myös semanttinen heräte-efekti (ks. *heräteasetelma* yllä): jos herätesana liittyy kriittiseen eli kohdesanaan (esim. *mehiläinen* ja *hunaja*), jos herätesana toistetaan (*hunaja* ja *hunaja*) tai jos sanat liittyvät toisiinsa semanttisesti (*sokeri* ja *hunaja*), sanan N400 on pienempi.

Syntaktisen prosessoinnin ensimmäisen osan (engl. *first pass*) ajatellaan olevan automaattista. Tätä automaattista, varhaista prosessointia heijastaa **ELAN**-komponentti eli aikainen vasemman etulohkon negatiivisuushuippu (engl. *early left-anterior negativity*) noin 100–200 ms kriittisen sanan esittämisen jälkeen. Varhaisen prosessoinnin ajatellaan perustuvan sanaluokkien erotteluun, taivutuksen tunnistamiseen ja lauserakenteen alustavaan hahmottamiseen. ELAN-efektiä voidaan tarkastella esimerkiksi seuraavanlaisten lauseiden avulla:

*(2) The goose was in the *fed.*
*The goose was in the barn fed.*

**P600**-komponentti liitetään myös syntaktiseen prosessointiin. Se on positiivinen aalto noin 600 ms kriittisen sanan esittämisen jälkeen. P600-komponentti heijastaa siis reaktiota syntaktiseen poikkeamaan. Useimmat tutkijat esittävät, että P600-efekti liittyy lauseen uudelleenanalysointiin. Esimerkiksi seuraavanlaisten lauseiden avulla voidaan tarkastella P600-efektiä:

*(3) Pikku poika *pelata pallolla.*
*Pikku poika pelaa pallolla.*

Kokoavasti voidaan todeta, että offline-menetelmiä, kuten nimeäminen ja heräteasetelma, käytetään kielenprosessoinnin ei-ajalliseen tarkasteluun. Usein näitä menetelmiä käytetään, kun pilotoidaan online-kokeissa käytettävää materiaalia. Online-menetelmillä saadaan puolestaan ajallista tietoa kielenprosessoinnista. Reaktioaikaan perustuvilla menetelmillä saadaan tietoa yhdestä mittauspisteestä, esimerkiksi yhden tehtävän suorittamiseen mennyt aika. Jatkuvaan mittaukseen perustuvilla menetelmillä voidaan tarkastella KH:n kielenprosessointia ajassa etenevänä prosessina.

**Kokeellisen kielentutkimuksen vahvuuksia ja haasteita**

**Selkeä tutkimuskysymys**. Kun tutkimuskysymys on hyvin muodostettu, tutkimus etenee loogisesti tutkimusraporttiin asti.

**Yleistettävyys**. Hyvin tehdyn kokeellisen tutkimuksen tulokset ovat yleistettävissä, riippumatta taustateoriasta.

**Kiinnostavia tutkimuskohteita ja ryhmässä työskentelyä**. Kokeellinen kielentutkimus on jännittävää, kiinnostavaa ja mukaansatempaavaa. Aiemmat tutkimukset nostavat esiin entistäkin kiinnostavampia teemoja tutkittavaksi. Kokeellista tutkimusta tehdään useimmiten ryhmässä, ja tutkimusta tehdessä oppii muilta.

**Datan määrä**. Aineiston suuri määrä mahdollistaa tutkimuskysymyksen monitasoisen tarkastelun. Erityisesti jatkuvan mittauksen menetelmistä saadaan erittäin paljon dataa, mikä toisaalta edellyttää tutkijalta valintoja tarkastelutavan suhteen. Tarkasti määritelty tutkimuskysymys auttaa näiden valintojen tekemisessä.

**Fysiologiset tekijät**. Ärsykkeet tulee valita ja asemoida koeasetelmiin niin, että ne eivät vaikuta tuloksiin. Esimerkiksi silmänliikemittauksissa tulee varmistaa, että tarkasteltavat sanat eivät ole rivin alussa tai lopussa, koska näiden alueiden lukeminen on erilaista kuin rivin keskiosan. EEG-mittauksessa puolestaan tarkastellaan kohdesanaa usein yksittäisenä tai lauseen viimeisenä tai toiseksi viimeisenä sanana, jotta EEG-rekisteröinti voidaan lukita näihin sanoihin.

**Psykologis-semanttiset tekijät**. Kielellisten ärsykkeiden valinnassa tulee ottaa huomioon esimerkiksi kohdesanan elollisuus, koska tällaiset psykologis-semanttiset tekijät voivat vaikuttaa KH:iden kielenprosessointiin ja sitä kautta tuloksiin. Psykologis-semanttiset tekijät voivat joskus olla myös tutkimuksen tarkastelun kohteena.

**Määrälliset tekijät**. Kun valitaan kielellisiä ärsykkeitä koeasetelmaan, on tärkeää ottaa huomioon myös kohdesanojen määrälliset tekijät, esimerkiksi sanan pituus ja yleisyys. Usein lyhyet sanat ovat myös yleisiä, minkä

vuoksi voi olla vaikeaa erottaa, kumpi muuttuja koeasetelmassa vaikuttaa tarkasteltavaan ilmiöön.

**Laadulliset tekijät**. Koeasetelmassa tulee pyrkiä kontrolloimaan tarkasti kohdesanojen sanaluokat, vaikka näiden rajat ovat joskus liukuvia (esim. *lukea* [verbi] vs. *lukeminen* [verbaalisubstantiivi]). On myös tärkeää kontrolloida KH:iden sukupuoli, ikä, koulutus ja mahdolliset muut taustamuuttujat.

**Tutkimuksen työläys ja kalleus**. Koeasetelmat ovat usein vaativia ja data on sotkuista, joten tutkimustulosten saaminen vaatii runsaasti suunnittelua, työtä ja aikaa. Lisäksi MRI- ja MEG-laitteiden käyttö saattaa olla maksullista.

## 3.4. Psykolingvistinen esimerkkitutkimus

Tässä luvussa esitellään psykolingvistinen esimerkkitutkimus visuaalisesta sanantunnistuksesta ("L1 and L2 word recognition in Finnish", Vainio, Pajunen & Hyönä 2014). Tutkimuksessa tarkasteltiin, poikkeaako venäjänkielisten ja kiinankielisten aikuisten suomenoppijoiden suomen kielen sanantunnistus äidinkielisten suomenkielisten sanantunnistuksesta ja toisistaan. Tutkimuksen tarkoituksena oli selvittää, missä määrin äidinkielen ja kohdekielen erilaisuus vaikuttaa sanantunnistukseen kohdekielellä. Riippumattomina muuttujina olivat suomen kielen morfologinen rakenne ja sanan kannan läpinäkyvyys. Nämä valittiin tarkastelun kohteeksi, sillä suomen kielessä sanan kantaan lisättävät johtimet ja taivutusmuodot aiheuttavat usein morfofonologisia muutoksia sanavartaloon, esim. *onki* -> *onge-ssa*. Tutkittavat substantiivit olivat yleisiä ja lyhyitä, ja ne esitettiin yksikön nominatiivissa, partitiivissa tai genetiivissä. Kontrollitilanteena oli nominatiivi (*koulu*), puhtaasti agglutinatiivista sijataivutusta edusti partitiivi (*tuolia*), ja sanoja, joissa oli taivutuksen lisäksi sanavartalon vaihtelua, edusti genetiivi (*kengän*). Kiinalaiset ja venäläiset olivat suomen kielen taidoltaan B2-tasolla eli itsenäisiä kielenkäyttäjiä. Taulukossa 2 havainnollistetaan tämän tutkimuksen kannalta keskeisiä kielten eroja.

| suure | suomi | venäjä | kiina |
|---|---|---|---|
| maksimi kategoriamäärä per sana | 1 + 4 | 1 + 4 | 1 + 0 |
| keskimääräinen kategoriamäärä per sana | 1 + 1 | 1 + 1 | 1 + 0 |
| alisteisia morfeemeja | kyllä | kyllä | ei |
| semanttiset roolit sijamerkitty | kyllä | kyllä | ei |
| kannan transparenssin vaihtelua | kyllä | kyllä | ei |
| kirjoitustapa | segmentaalinen (latinalainen) | segmentaalinen (kyrillinen) | logografinen (kiinalainen) |
| kategoriamäärä per sana = yhteenlaskettu morfologisten luokkien määrä sanassa (esim. *talo-i-ssa-mme-kin* = 1 + 4) | | | |

Taulukko 2. Suomen, venäjän ja kiinan typologisia piirteitä.

Taulukosta 2 selviää, että kiinankielisillä ei ole äidinkielessään kokemusta sijataivutuksesta mutta venäjänkielisillä on. Lisäksi venäjässä on genetiivin taivutuksen yhteydessä kannan vaihtelua, kun taas kiinan kielessä ei tällaista ilmiötä ole. Mainittakoon, että kohdesanat vertaistettiin lemmafrekvenssin (sanan kaikkien sijamuotojen yhteenlasketun summan), pintafrekvenssin (sanan tietyn sijamuodon frekvenssin), sanapituuden (kirjainmäärän), konkreettisuusarvioinnin ja omaksumisiän arvioinnin mukaan. Sanat olivat yleisiä, koska tarkoitus oli, että keskitasoiset suomenoppijat tietäisivät sanat; tyypillisesti vieraan kielen prosessointia tutkitaan melko sujuvasti kieltä osaavilla KH:illä.

Kokonaisuutena tutkimuksessa käytettiin kahta menetelmää: visuaalista sanantunnistusta ja silmänliikkeiden rekisteröintiä KH:iden lukiessa lauseita. Menetelmillä tarkasteltiin, vaikuttaako äidinkieli suomen kielen prosessointiin sanatasolla ja lausetasolla. Mikäli näin olisi, morfologinen monimutkaisuus vaikuttaisi eri tavalla kiinalaisiin ja venäläisiin sanantunnistuksessa, mutta sitä vastoin määrite-pääsanakongruenssia prosessoitaessa kiinalaisten ja venäläisten tulosten pitäisi

olla samankaltaisia. Aiempiin tutkimuksiin perustuvat mallit edustavat vastakkaisia näkemyksiä siitä, vaikuttaako äidinkielen ja kohdekielen erilaisuus oppimiseen. Lisäksi on esitetty, että morfofonologinen vaihtelu ylipäätään vaikeuttaa vieraan kielen rakenteiden oppimista (DeKeyser 2005).

**Kilpailumalli** (engl. *Unified Competion Model*, MacWhinney 2005) ennustaa, että kielenoppijat todennäköisesti käyttävät äidinkielensä rakenteita uuden kielen oppimisessa: oppiminen on helpompaa, jos kielenoppijan äidinkielessä on samankaltaisia rakenteita kuin opittavassa kielessä. Tämän tutkimuksen osalta kilpailumalli olettaa, että venäläisille sijataivutus ei ole merkittävä ongelma, joten he prosessoivat sanoja samankaltaisesti kuin suomalaiset.

Pienemann (1998) on puolestaan esittänyt **prosessoitavuusteorian** (engl. *Processability Theory*), jossa on kaksi keskeistä periaatetta. Ensimmäisen periaatteen mukaan äidinkielen omaksuminen perustuu tiettyyn poikkeuksettomaan kehitykselliseen järjestykseen. Toinen keskeinen periaate on, että tämä oppimishierarkia pätee myös vieraan kielen oppimiseen, eli äidinkielen ja kohdekielen erolla ei ole merkitystä. Näin ollen prosessoitavuusteoria olettaa venäläisten ja kiinalaisten suoriutuvan suomen sanantunnistuksesta samalla tavalla.

Kolmas näkökulma teemaan on **morfologisen jäsennyskyvyn puutteellisuuden hypoteesi** (engl. *Morphological Decomposition Deficiency Hypothesis*), joka perustuu ajatukseen siitä, että vieraan kielen (L2) oppijat eivät pysty prosessoimaan vieraan kielen taivutusmorfologiaa. Taustalla on kaksi näkemystä, joista ensimmäinen perustuu aivokuvantamisessa havaittuihin tuloksiin (Ullman 2001). Niiden mukaan L2-oppijat prosessoivat leksikkoa ja semantiikkaa samalla muistialueella kuin äidinkieliset. VanPatten (2004) puolestaan esittää mallin, jossa keskeisenä ajatuksena on merkityksen ensisijaisuus, mikä tarkoittaa, että kielenoppijat oppivat ensin prosessoimaan sanojen merkityksen ennen kuin he oppivat prosessoimaan sanojen muodon. Siten oppijat käyttävät mieluummin leksikaalista kuin morfologista tietoa päättäessään lauserakenteen ajallisesta viittauksesta (VanPatten & Jegerski 2010).

Kaikille yllämainituille malleille on yhteistä, että niiden mukaan morfologisesti mutkikkaat taivutetut sanat ovat jäsennettyinä

mentaalileksikossa. Vaihtoehtoinen tulkinta, esim. **yhteneväisyysteoria** (engl. *Convergent Theory*, Gonnerman, Seidenberg & Andersen 2007), olettaa, että taivutusmorfologia ei ole mentaalileksikossa erillisinä kantoina ja affikseina. Jos näin on, morfologisen monimutkaisuuden ei pitäisi vaikuttaa reaktioaikoihin.

Esiteltävän tutkimuksen tulokset ovat alla olevassa taulukossa 3. Tilastoanalyysit osoittivat, että reaktioaikoja tarkasteltaessa saatiin tilastollisesti merkitsevät sijamuodon ja äidinkielen (L1) päävaikutukset sekä näiden yhdysvaikutus. Avattuna yhdysvaikutus paljasti, että äidinkieliset suomenkieliset prosessoivat nominatiivit merkitsevästi nopeammin kuin taivutetut sijat (partitiivi ja genetiivi), joiden välillä ei ollut merkitsevää eroa. Tämän tutkimuksen reaktioaikatulos suomalaisten osalta voidaan siis esittää lyhyesti seuraavasti: nominatiivi < partitiivi = genetiivi. Tulokset toistivat aiempien suomen kielen tutkimusten tulokset.

| | | nominatiivi | | partitiivi | | genetiivi | |
|---|---|---|---|---|---|---|---|
| L1 | tarkasteltava tekijä | KA | KH | KA | KH | KA | KH |
| suomen-kieliset | reaktioaika | 665 | 130 | 702 | 121 | 719 | 130 |
| | virhemäärä | 1,3 | 2,4 | 3,8 | 4,7 | 5,3 | 5,8 |
| kiinan-kieliset | reaktioaika | 927 | 192 | 961 | 215 | 1078 | 267 |
| | virhemäärä | 10,5 | 5,1 | 11,4 | 10,1 | 29,7 | 17,9 |
| venäjän-kieliset | reaktioaika | 880 | 175 | 1035 | 222 | 1134 | 255 |
| | virhemäärä | 8,8 | 15,6 | 8,8 | 8,1 | 18,3 | 8,7 |

KA = keskiarvo; KH = keskihajonta.

Taulukko 3. Keskimääräiset reaktioajat (millisekunteina) ja virhemäärät (prosentteina) sekä niiden keskihajonnat suomalaisilla, kiinalaisilla ja venäläisillä.

Kiinankielisten reaktioaikojen tilastotulokset osoittivat, että kiinankielisillä genetiivin prosessointi poikkesi merkitsevästi sekä nominatiivin että partitiivin prosessoinnista mutta nominatiivin ja partitiivin prosessointi oli samankaltaista. Tulos viittaa siihen, että kiinalaiset eivät

prosessoineet morfologista rakennetta, sillä partitiivin ja nominatiivin välillä ei ollut eroa. Sen sijaan sanavartalon läpinäkyvyys vaikutti prosessointiin siten, että kun sanavartalossa oli vaihteluita, niin tunnistusaika oli merkitsevästi pidempi. Tulokset voidaan esittää lyhyesti seuraavasti: nominatiivi = partitiivi < genetiivi.

Venäjänkielisillä nominatiivien prosessointi oli merkitsevästi nopeampaa kuin partitiivien ja genetiivien. Myös partitiivit prosessoitiin nopeammin kuin genetiivit, vaikkei tämä tulos ollut kokonaisuudessaan merkitsevä. Tulos viittasi siihen, että venäjänkielisten prosessointi oli melko samanlaista kuin suomenkielisten eli lyhyesti: nominatiivi < partitiivi < genetiivi.

Kun tarkasteltiin virheitä eli tilanteita, joissa KH:t olivat vastanneet väärin, tilastotulokset osoittivat, että erikieliset puhujaryhmät erosivat toisistaan merkitsevästi ja myös eri sijoissa tehtiin tilastollisesti merkitsevästi eri määriä virheitä. Äidinkielen ja sijan yhdysvaikutus oli myös tilastollisesti merkitsevä, eli suomenkieliset tuottivat kaikki sijat samalla tavalla oikein. Sekä kiinankielisillä että venäjänkielisillä virhemäärät olivat samankaltaiset, eli genetiivissä tehtiin enemmän virheitä kuin nominatiivissa ja partitiivissa.

Tulokset tukevat kilpailumallin ennustetta, sillä venäjänkieliset prosessoivat suomen sanoja samankaltaisesti kuin suomenkieliset mutta kiinankielisillä vain sanavartalon läpinäkyvyys vaikutti prosessointiin. Sen sijaan prosessoitavuusteoria ja morfologisen jäsennyksen puutteellisuusteoria eivät saa vahvistusta. Virheiden tarkastelu osoitti selvästi, että sanavartalon läpinäkyvyys on merkittävä tekijä L2-oppijoiden prosessoinnissa, kuten DeKeyser (2005) ennustaa.

Yllä esitettyjen tulosten perusteella voisi olettaa, että koska kiinankieliset suomen oppijat eivät kiinnitä taivutusmorfologiaan huomiota sanantunnistuksessa, he eivät myöskään hyödynnä taivutusmorfologiaan perustuvaa määritteen ja pääsanan kongruenssia lauseen lukemisessa. Samat kiinankieliset KH:t osallistuivat myös lukemiskokeeseen (Vainio, Pajunen & Hyönä 2016), jossa sekä kiinankieliset että venäjänkieliset hyödynsivät suomen määritteen ja pääsanan kongruenssia samanlaisesti. Tulos on selitettävissä sillä, että kiinankielisillä on äidinkielensä perusteella kokemusta siitä, että lausekkeen osia liitetään yhteen eri

keinoin, vaikkei heillä ole kokemusta sanan sijataivutuksesta sinänsä. Tällainen ajatus sopii hyvin esim. Sabourinin ja Stowen (2008) malliin, jossa korostetaan prosessointirutiinien siirtoa (engl. *Transfer of Processing Routines*) representaatioiden siirron sijaan, eli omassa kielessä oleva sanojen yhdistämisen periaate voidaan yleistää toisen kielen oppimiseen.

## 3.5. Neurolingvistinen esimerkkitutkimus

Tässä luvussa esitellään yksi suomalainen neurolingvistinen väitöskirjatutkimus kielitieteen alalta. Tutkimus käsittelee MS-tautia sairastavien kieltä ja sen suhdetta taudin yhteen oireeseen, kognitiiviseen uupumukseen (Heikkola 2017). Tutkimuksessa tarkasteltiin, eroavatko MS-tautia sairastavien kielelliset kyvyt ja puheentuotto terveistä verrokeista ja korreloiko MS-taudin yksi oire, kognitiivinen uupumus, puheentuoton kanssa. *Kognitiivisella uupumuksella* tarkoitetaan subjektiivista energian puutetta, jonka yksilö tai hänen hoitajansa kokee häiritsevän tavanomaisia tai toivottuja toimintoja. Koska kielellinen toiminta on keskeinen MS-tautia sairastavien arkielämässä selviytymisessä (Murdoch & Theodoros 2000), oli tärkeää selvittää, millainen vaikutus kognitiivisella uupumuksella on kieleen. MS-tautia sairastavien kielihäiriöitä ei ole aiemmin juurikaan tutkittu kielitieteellisestä näkökulmasta, eikä kognitiivisen uupumuksen vaikutusta MS-tautia sairastavien kieleen ole aiemmin tutkittu lainkaan.

Koska MS-taudin oireet vaihtelevat suuresti yksilöiden välillä, myös siihen liittyvien kielihäiriöiden kirjo on suuri. Tapaustutkimuksissa on todettu, että MS-tautiin liittyy ongelmia nimeämisessä ja toistamisessa (Olmos-Lau, Ginsberg & Geller 1977). Ryhmätutkimuksissa on lisäksi esitetty, että myös puheen sujuvuudessa ja lukemisessa sekä kirjoittamisessa voi olla ongelmia (Murdoch & Lethlean 2000). 35 tutkimusta kattavassa metatutkimuksessa on todettu, että MS-tautia sairastavilla muutokset sanasujuvuudessa ovat yleisimpiä kielen häiriöistä (Henry & Beatty 2006). MS-tautia sairastavien kieltä ei kuitenkaan olla juurikaan tarkasteltu teksti- tai diskurssitasolla (Arrondo, Sepulcre, Duque, Toledo

& Villoslada 2010). Nykyään erilaisia neurologisperäisiin sairauksiin liittyviä kielihäiriöitä on alettu tarkastella semispontaanien narratiivien eli vapaasti kuvamateriaalin perusteella kerrottujen tarinoiden kieltä analysoimalla (ks. Lind, Kristoffersen, Moen & Simonsen 2009).

Tutkimuksessa riippumattomina muuttujina olivat MS-tautidiagnoosi sekä MS-taudin tyyppi ja vakavuusaste. MS-tutkimuksen aineisto oli kolmitahoinen. MS-tautidiagnoosin sekä taudin tyypin ja vakavuusasteen määritteli neurologi. Toiseksi KH:t testattiin laajalla neuropsykologisella testipatteristolla yhteistyössä neuropsykologin kanssa, jotta saatiin selville KH:iden kognitiivinen perustaso sekä kielellisen kognition taso. Lisäksi arvioitiin sekä koetun että mitattavissa olevan kognitiivisen uupumuksen määrää. Kolmanneksi KH:t osallistuivat kielelliseen testaukseen, jossa he kertoivat semispontaaneja tarinoita kuvakirjojen pohjalta. Kuvakirjojen välissä oli tehtävän vaikeuttamiseksi kaksi täytetehtävää, joissa KH:t puhuivat myös kuvamateriaalin pohjalta. Kielellinen testaus oli pitkäkestoinen, koska tutkimuksessa haluttiin tarkastella kognitiivisen uupumuksen vaikutusta kieleen. Aineistoa tarkasteltiin tilastollisesti ANOVA-testien ja korrelaatiokerrointen avulla.

Aiemman tutkimuksen perusteella oletettiin, että MS-tautia sairastavat suoriutuvat heikommin kognitiota arvioivista tehtävistä ja että he kokevat enemmän kognitiivista uupumusta. Lisäksi oletettiin, että MS-tautia sairastavien sujuvuus tarinoiden kertomisessa on heikompi kuin verrokeilla. Oletettiin myös, että MS-tautia sairastavilla suurempi kognitiivinen uupumus korreloi tarinoiden sujuvuuden heikkenemisen kanssa, toisin kuin verrokeilla. Lisäksi oletettiin, että vakavampaa MS-taudin muotoa sairastavilla kognition ja kielen häiriöt olisivat vakavampia kuin lievempää taudin muotoa sairastavilla.

Tutkimuksen tulosten mukaan MS-tautia sairastavien kognitiivinen perustaso oli heikentynyt verrokkeihin verrattuna. Tämä on linjassa aiemman tutkimuksen kanssa (Kujala, Portin, Revonsuo & Ruutiainen 1994). Yllättävästi MS-ryhmän kielellinen kognitio oli osittain säilynyt, esimerkiksi nimeämisen osalta, vaikka juuri nimeämisen on aiemmissa tutkimuksessa osoitettu heikentyneen MS-taudissa (Murdoch & Lethlean 2000; Kujala, Portin & Ruutiainen 1996). Sekä subjektiivista että objektiivista kognitiivista uupumusta esiintyi enemmän

MS-ryhmässä. Oletuksen vastaisesti lievempää MS-taudin muotoa sairastavilla oli enemmän puutteita kognitiivisissa kyvyissä.

Myös kielen suhteen MS-ryhmä suoriutui heikommin kuin verrokkiryhmä. MS-tautia sairastavien tarinat eivät olleet yhtä sujuvia kuin verrokeilla. Kognitiivisen uupumuksen vaikutus näkyi MS-ryhmällä siten, että tarinoiden alut olivat sujuvampia kuin niiden loput. Lisäksi MS-tautia sairastavien ensimmäiset tarinat olivat sujuvampia kuin heidän kertomansa kolmannet, kielitestauksen viimeiset tarinat. Yhden tarinan sisällön tarkastelussa MS-tautia sairastavilla ilmeni enemmän täysin virheellisiä sisältöyksikköjä, mutta muuten tarinoiden sisällöissä ei ollut laadullisia eroja ryhmien välillä. Oletuksen mukaisesti lievempää MS-taudin muotoa sairastavilla tarinoiden sujuvuus ja sisältöjen koherenssi olivat säilyneet paremmin kuin taudin vakavampaa muotoa sairastavilla.

MS-tautia sairastavilla oli myös yhteys objektiivisesti mitatun kognitiivisen uupumuksen sekä kolmen suullisesti kerrotun tarinan keston välillä, mikä oli hypoteesin mukaista. Lisäksi kielellisen testauksen aikana arvioidun puheentuoton aiheuttaman koetun väsymyksen ja tarinoiden sisällön koherenssin välillä oli yhteys. Oletetusti myös tarinoiden sujuvuuden ja neuropsykologisella testillä arvioidun semanttisen sanasujuvuuden (eli esimerkiksi kuinka monta eläintä KH ehtii nimetä minuutissa) välillä oli yhteys. Niin ikään oletetusti vakavampaa MS-taudin muotoa sairastavilla tämä yhteys oli selvempi kuin taudin lievempää muotoa sairastavilla.

Tutkimuksen tulosten perusteella MS-tautia sairastavien kognitiivinen perustaso ja kielelliset kyvyt ovat selvästi heikentyneet terveisiin verrokkeihin verrattuna. Sekä koetun että objektiivisesti mitatun kognitiivisen uupumuksen ja kielentuottamisen välillä oli yhteyksiä: korkeampi kognitiivinen uupumus näkyi usein kielentuotossa vähentyneenä sujuvuutena tai tarinan sisällön koherenssin puutteina.

## 3.6. Kokeellisen kielentutkimuksen lähialat

Kuten edellä on esitetty, psyko- ja neurolingvistiikka ovat hyvin poikkitieteellisiä aloja. Psyko- ja neurolingvistiikka ovat ottaneet vaikutteita monilta eri aloilta, joista seuraavaksi esitellään tärkeimpiä.

**Yleinen kielitiede**. Yksi suurimmista kielitieteellisistä vaikutteista kokeelliseen kielentutkimukseen on strukturalismi (de Saussure 1916), jonka mukaan kieli koostuu sen eri aspekteja esittävistä järjestelmistä (fonologia, morfologia, semantiikka, syntaksi ja leksikko). Myös Noam Chomskyn kehittämä generatiivinen kielioppi on vaikuttanut erityisesti syntaksin prosessointiin liittyvään kokeelliseen kielentutkimukseen. Lisäksi kognitiivinen semantiikka eli yritys tehdä mentaalimalleja semanttisista prosesseista on vaikuttanut jonkin verran esimerkiksi afasiassa esiintyvien semanttisten ongelmien tarkasteluun. Nykyään myös funktionalismi eli kielen funktioihin perustuva näkemys kielestä vaikuttaa kokeelliseen kielentutkimukseen. Kokeellisen kielentutkimuksen taustalla vaikuttavat usein myös kognitiivisen kielitieteen näkemykset (ks. Kieliopin tutkimus ja kielioppiteoriat tk.).

**Kliininen kielitiede, aivoperäisten kielihäiriöiden puhe- ja kieliterapia**. Monet afasiaterapian arviointi- ja kuntoutusmenetelmät perustuvat klassisiin viitekehyksiin, kuten Lurian teoriaan, joiden mukaan toiminnallisia järjestelmiä kuntoutetaan askeleittain. Muut terapiat liittyvät Wernicke–Lichtheim–Geschwind-viitekehyksen teoreettisiin oletuksiin. Näissä terapioissa keskitytään aktivoimaan säilyneitä aivojen keskuksia, joita ei primaaristi käytetä kielen toimintoihin, jotta nämä keskukset voisivat välittää kielitoimintoja. Näitä terapiamuotoja ovat mm. VIC (engl. *visual communication therapy*, Gardner, Zurif, Berry & Baker 1976), VAT (engl. *visual activation therapy*, Helm-Estabrooks, Fitzpatrick & Barresi 1982) ja MIT (engl. *melodic intonation therapy*, Albert, Goodglass & Helm 1973). Jotkin uudemmat terapiamuodot ovat kommunikaatiopohjaisia, kuten PACE (engl. *Promoting Aphasics' Communicative Efficiency*, Carlomagno 1994). Nykyään käytettäviin terapiamuotoihin vaikuttavat myös pragmaattinen koulukunta, neurolingvistinen koulukunta ja kognitiivisen neuropsykologian koulukunta.

**Aivokuvantaminen ja aivojen dynaamisten prosessien tutkimus**. Aivokuvantamisen kehitys on ollut suuri läpimurto kokeellisessa kielentutkimuksessa. Aivorakenteen kuvantaminen tietokonetomografialla ja MR-kuvantaminen sekä aivotoiminnan kuvantaminen PET:llä, fMRI:llä, MEG:llä, EEG:llä ja herätevasteilla mahdollistavat kielellisen ja kognitiivisen prosessoinnin tarkastelun reaaliajassa sekä terveissä että vaurioituneissa aivoissa. Aivokuvantamisen ja psykolingvististen kokeellisten tekniikoiden yhdistäminen ei välttämättä ole suoraviivaista eikä tulosten tulkinta yksinkertaista, mutta nämä yhdistämällä saadaan uutta ja tärkeää tietoa kielen prosessoinnista.

**Psykologia ja kognitiotiede**. Muun muassa Weisenburgin ja McBriden (1935) käyttämillä psykologian arviointimenetelmillä, joilla on tarkasteltu kielellistä ja ei-kielellistä kognitiota afasiaa sairastavilla ja kontrolliryhmillä, on ollut laaja vaikutus erityisesti neurolingvistiikan menetelmiin. Kokeellisessa kielentutkimuksessa käytetään edelleen kokeellisia asetelmia ja tilastoanalyyseja, joita käytettiin 1960- ja 1970-luvuilla afasian ryhmätutkimuksissa. 1980-luvulta lähtien afasiaa on tarkasteltu yhä enemmän myös kokeellisten yksittäistapausten kautta (näiden tutkimustapojen käytöstä kokeellisessa kielentutkimuksessa ks. Caramazza 1984).

**Tietokonesimulaatiot ja mallintaminen**. Tietokonesimulaatioilla voidaan testata ja hienosäätää kielen prosessointimalleja. Simulaatioilla voidaan esimerkiksi tutkia, miten hypoteettiset leesiot eli tietyn aivoalueen vammat vaikuttavat kognitiiviseen prosessointiin hypoteettisten mallien mukaan. Kielen tuottamisen ja ymmärtämisen seriaaliprosessointimalleja on käytetty kokeellisessa kielentutkimuksessa jo 1980-luvulta. Viime vuosikymmeninä niitä on käytetty viitekehyksinä useimmissa psyko- ja neurolingvistiikan kokeellisissa tutkimuksissa sekä terveillä että aivovaurion saaneilla KH:illa. Viime vuosina erityisesti rinnakkaisprosessointimallien ja konnektionististen eli keinoneuraaliverkostojen simulaatioiden käyttö on lisääntynyt kokeellisessa kielentutkimuksessa.

# 4. Yhteenveto

Tässä luvussa on esitelty lyhyesti psyko- ja neurolingvistiikan tutkimusta. Ala on ollut olemassa nykyisessä muodossaan 1900-luvun lopulta, ja erityisesti tietokoneiden ja aivokuvantamisen valtava kehitys on avannut uusia mahdollisuuksia tarkastella mielen, aivojen ja kielen yhteyksiä.

Kokeellinen tutkimus on useimmiten ryhmätyötä, johon osallistuvat eri alojen ammattilaiset. Koeasetelmien suunnittelu ja kokeiden valmistelu on aikaa vievää, ja usein on vaikeaa kontrolloida kaikkea, mikä voi vaikuttaa tutkimuksen tuloksiin. Hyvällä koesuunnittelulla ja järjestelmällisellä testaamisella voidaan kuitenkin saavuttaa tuloksia, jotka kertovat jotain uutta siitä, miten me ihmiset prosessoimme kieltä.

Psyko- ja neurolingvistiikasta kiinnostuneen on hyvä tutustua omatoimisesti lähitieteisiin, erityisesti yleiseen kielitieteeseen, psykologiaan ja kognitiiviseen neurotieteeseen. Myös tutkimustekniikkaan, esimerkiksi silmänliikekameroihin ja EEG-laitteisiin, on hyvä tutustua. Parhaat mahdollisuudet psyko- ja neurolingvistisen tutkimuksen tekemiseen ovat tutkimusryhmässä, jossa kokeneempien tutkijoiden johdolla pääsee tutustumaan mielen, aivojen ja kielen ihmeellisiin yhteyksiin.

**Aiheesta lisää:**


Grosjean, Francois & Li, Ping. 2012. *The Psycholinguistics of Bilingualism*. Somerset: Wiley-Blackwell.

Paradis, Michel. 2004. *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins.

Stemmer, Brigitte & Whitaker, Harry (toim.). 1998. *Handbook of Neurolinguistics*. San Diego: Academic Press.

Traxler, Matthew. 2012. *Introduction to Psycholinguistics: Understanding Language Science*. Chichester, West Sussex & Malden (MA): Wiley-Blackwell.

## KIRJALLISUUS


Ahlsén, Elisabeth. 2006. *Introduction to Neurolinguistics*. Amsterdam: John Benjamins.

Albert, Martin, Sparks, Robert W. & Helm, Nancy A. 1973. Melodic intonation therapy for aphasia. *Archives of Neurology* 29, 130–131.

Arrondo, Gonzalo, Sepulcre, Jorge, Duque, Beatriz, Toledo, Jon & Villoslada, Pablo. 2010. Narrative speech is impaired in multiple sclerosis. *European Neurological Journal* 2, 11–18. [verkkoaineisto]. [viitattu 7.1.2017]. Saatavissa: https://www.researchgate.net/publication/45825431_Narrative_Speech_is_Impaired_in_Multiple_Sclerosis.

Beeson, Pélagie M. & Robey, Randall R. 2006. Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychol. Rev.* 16:4, 161–169.

Byiers, Breanne J., Reichle, Joe & Symons, Frank J. 2012. Single-subject experimental design for evidence-based practice. *American Journal of Speech-Language Pathology* 21, 397–414.

Caramazza, Alfonso. 1984. The logic of neuropsychological research and the problem of patient classification in aphasia. *Brain and Language* 21, 9–20.

Carlomagno, Sergio. 1994. *Pragmatic Approaches to Aphasia (Promoting Aphasics' Communicative Efficiency)*. London: Singular.

DeKeyser, Robert. 2005. What makes learning second-language grammar difficult? A review of issues. *Language Learning* 55:S1, 1–25.

Gardner, Howard, Zurif, Edgar B., Berry, Thomas & Baker, Errol. 1976. Visual communication in aphasia. *Neuropsychologia* 14, 275–292.

Gonnerman, Laura, Seidenberg, Mark, & Andersen, Elaine. 2007. Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General* 136:2, 323–345.

Grönroos, Matti. 2008. *Johdatus tilastotieteeseen*. Helsinki: Finn Lectura.

Hampton, James. 1998. Between-subjects versus within-subject designs. Julkaisussa: Altmann, Gerry & Gathercole, Susan (toim.) *Laboratory Psychology: A Beginner's Guide*. Hove: Psychology Press Ltd, 15–38.

Heikkola, Leena Maria. 2017. *Kognitiivinen uupumus ja kieli MS-taudissa. Semispontaanien narratiivien määrällinen ja laadullinen analyysi*. Väitöskirja. Turku: Åbo Akademi.

Hellier, Elizabeth. 1998. Within-subject designs. Julkaisussa: Altmann, Gerry & Gathercole, Susan (toim.) *Laboratory Psychology: A Beginner's Guide*. Hove: Psychology Press Ltd, 39–58.

Helm-Estabrooks, Nancy A., Fitzpatrick, Patricia M. & Barresi, Barry J. 1982. Visual action therapy for global aphasia. *Journal of Speech and Hearing Disorders* 47:4, 385–389.

Henry, Julie D. & Beatty, William W. 2006. Verbal fluency deficits in multiple sclerosis. *Neuropsychologia* 44:7, 1166–1174.

Just, Marcel & Carpenter, Patricia. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review* 87:4, 329–354.

Klippi, Anu, Korpijaakko-Huuhka, Anna-Maija & Lehtihalmes, Matti. 2017. *Afasia: Aikuisiän kielihäiriöiden perusta ja kuntoutus*. Helsinki: Gaudeamus.

Kujala, Päivi, Portin, Raija, Revonsuo, Antti & Ruutiainen, Juhani. 1994. Automatic and controlled information processing in multiple sclerosis. *Brain* 117:5, 1115–1126.

Kujala, Päivi, Portin, Raija & Ruutiainen, Juhani. 1996. Language functions in incipient cognitive decline in multiple sclerosis. *Journal of the Neurological Sciences* 141:1–2, 79–86.
Kutas, Marta & Hillyard, Steven A. 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207:4427, 203–205.
Levelt, Willem. 2014. *A History of Psycholinguistics: The Pre-Chomskyan Era*. Oxford: Oxford University Press.
Lind, Marianne, Kristoffersen, Kristian E., Moen, Inger & Simonsen, Hanne G. 2009. Semi-spontaneous oral text production: measurements in clinical practice. *Clinical Linguistics and Phonetics* 23:12, 872–886.
MacWhinney, Brian. 2005. A unified model of language acquisition. Julkaisussa: Kroll, Judith & de Groot, Annette (toim.) *Handbook of Bilingualism: Psycholinguistic Approaches*. Oxford: Oxford University Press, 49–67.
McLeod, Saul. 2008. *Wilhelm Wundt*. [verkkoaineisto]. [viitattu 18.4.2018]. Saatavissa: https://www.simplypsychology.org/wundt.html.
Murdoch, Bruce E. & Lethlean, Jennifer B. 2000. Language disorders in multiple sclerosis. Julkaisussa: Murdoch, Bruce & Theodoros, Deborah (toim.) *Speech and Language Disorders in Multiple Sclerosis*. London: Whurr Publishers.
Murdoch, Bruce & Theodoros, Deborah (toim.). 2000. *Speech and Language Disorders in Multiple Sclerosis*. London: Whurr Publishers.
Nicol, Janet, Forster, Kenneth & Veres, Csaba. 1997. Subject–verb agreement processes in comprehension. *Journal of Memory & Language* 36:4, 569–587.
Nummenmaa, Lauri. 2009. *Käyttäytymistieteiden tilastolliset menetelmät*. Helsinki: Sanoma Pro.
Obler, Loraine K. & Gjerlow, Kris. 1999. *Language and Brain*. Cambridge Approaches to Linguistics. Cambridge: Cambridge University Press.
Olmos-Lau, Nick, Ginsberg, Myron D. & Geller, Joan B. 1977. Aphasia in multiple sclerosis. *Neurology* 27:7, 623–626.
Pienemann, Manfred. 1998. *Language Processing and Second Language Development: Processability Theory*. Amsterdam: Benjamins.
Rayner, Keith & Duffy, Susan. 1986. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition* 14:3, 191–201.
Sabourin, Laura & Stowe, Laurie. 2008. Second language processing: When are L1 and L2 processed similarly. *Second Language Research* 24:3, 397–430.
Saussure, Ferdinand de. 1916. *Cours de linguistique génerale*. Paris: Payot.
Swaab, Tamara Y., Ledoux, Kerry, Camblin, Christine & Boudewyn, Megan A. 2011. Language-related ERP components. Julkaisussa: Kappenman, Emily S. & Luck, Steven J. (toim.) *The Oxford Handbook of Event-Related Potential Components*. Oxford: Oxford University Press, 397–440.
Ullman, Michael. 2001. The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition* 4:2, 105–122.
Vainio, Seppo. 2006. Lausetason kokeellisesta tutkimuksesta: tuloksia, haasteita ja tulevaisuuden näkymiä. *Puhe ja kieli* 26:1, 3–21.

Vainio, Seppo, Pajunen, Anneli & Hyönä, Jukka. 2014. L1 and L2 word recognition in Finnish: Examining L1 effects on L2 processing of morphological complexity and morphological transparency. *Studies in Second Language Acquisition* 36:1, 133–162.

Vainio, Seppo, Pajunen, Anneli & Hyönä, Jukka. 2016. Processing modifier-head agreement in L1 and L2 Finnish: An eye-tracking study. *Second Language Research* 32:1, 3–24.

VanPatten, Bill (toim.). 2004. *Processing Instruction: Theory, Research, and Commentary.* Mahwah (NJ): Erlbaum.

VanPatten, Bill & Jegerski, Jill (toim.). 2010. *Research in Second Language Processing and Parsing*. Amsterdam: John Benjamins.

Weisenburg, Theodore & McBride, Katherine. 1935. *Aphasia*. New York (NY): Hafner.

Wundt, Wilhelm. 1975. Völkerpsychologie. Vol. 2. Die Sprache. Oxford: David Brown Book Company.

# English abstract

*Methods in Linguistics* is a collection of articles presenting a broad variety of methods and approaches in the field of linguistics. It offers structure to the spectrum of methodological possibilities and helps the reader to identify the field of application as well as the strengths and weaknesses of different methods. The book consists of an extensive introductory part and a variety of articles, each written by experts of the method in question. It discusses questions related to different kinds of data and data collection, as well as methods used for analysing data. Since the methods used in linguistics are often related to a specific linguistic current, the book offers examples from a wide range of linguistic approaches.

The book is addressed to students, researchers and other readers interested in the methods used in linguistics. The general methodological and metatheoretical knowledge offered in the introduction guides the reader through the different methodological choices presented and helps the reader to select the best method to meet her needs. First, the research process is explained step-by-step from the selection of topic and data to questions related to research design, analysis and the reporting of the results. The second part of the introduction focuses on fundamental theoretical and methodological questions, such as the criteria for knowledge, the nature of scientific knowledge and the scientific method. The third part addresses more specifically the methodology of linguistics, discussing the multifaceted nature of natural language and linguistics as a discipline. The introduction also covers many current topics in science, such as research ethics, data protection and open science principles.

The book can be used as self-study material by students and researchers alike, or as course material in higher education. Learning is supported by the careful definition of terms, extensive indexes and additional readings suggested for each topic.

# Kirjoittajat

FT TERHI AINIALA on nimistöntutkimukseen erikoistunut suomen kielen yliopistonlehtori Helsingin yliopistossa. Hänen tutkimusalojaan ovat erityisesti sosio-onomastiikka ja kaupunkinimistön tutkimus.
https://orcid.org/0000-0003-4998-9983

FT EKATERINA GRUZDEVA on Helsingin yliopiston yleisen kielitieteen yliopistonlehtori. Hänen erikoisalanaan on etnolingvistiikka ja kielitieteellinen kenttätutkimus, erityisesti Venäjän Kaukoidässä puhuttu nivhin kieli. Vuosina 2019–2021 hän toimii tutkijana Helsingin yliopiston tutkijakollegiumissa.
https://orcid.org/0000-0001-8898-649X

FT MARKUS HAMUNEN on suomen kielen post doc -tutkija. Hänen tutkimusalojaan on kieliopin tutkimus kognitiivisen kielentutkimuksen viitekehyksessä. Hän on kiinnostunut myös fennistiikan ja yleisen kielitieteen oppihistoriasta sekä yleisesti kielitieteen metateoriasta.
https://orcid.org/0000-0003-2786-6392

FT KATARIINA HARJUNPÄÄ työskentelee tutkijatohtorina Helsingin yliopiston humanististen tieteiden ohjelmassa. Hänen alaansa ovat keskustelunanalyysi ja vuorovaikutuslingvistiikka, ja erityisesti häntä kiinnostavat multimodaalisuus, monikieliset käytänteet sekä osallistujuus ja toimijuus vuorovaikutuksessa.
https://orcid.org/0000-0002-4586-1563

FT LEENA MARIA HEIKKOLA on suomen kielen yliopisto-opettaja Åbo Akademissa. Hänen alaansa ovat toisen kielen oppiminen ja opettaminen, laulut kielenopetusmenetelmänä, suomi toisena kielenä, kieli- ja kulttuuritietoisuus, MS-tautiin liittyvät kielihäiriöt, afasiatutkimus sekä kliininen kielitiede.
https://orcid.org/0000-0002-1795-7656

FT PIRJO HIIDENMAA on tietokirjallisuuden ja tietokirjoittamisen tutkimuksen professori Helsingin yliopistossa. Hänen alaansa ovat tekstin- ja diskurssintutkimus, lukutaidon ulottuvuudet, tietokirjallisuus ja sen esitystavat sekä koulujen kirjallisuuskasvatus.
https://orcid.org/0000-0002-1925-8445

FT TUOMAS HUUMO on suomen kielen professori Turun yliopistossa ja vieraileva suomen kielen professori Tarton yliopistossa Virossa. Hänen keskeisiä tutkimusalojaan ovat kognitiivinen kielitiede, semantiikka ja kielioppi ja tutkimusaiheitaan kielioppiin koodautuneet metaforat, aspekti, sijat, adpositiot ja astemääritteet.
https://orcid.org/0000-0002-6589-0732

FT LOTTA JALAVA on suomalais-ugrilaisten kielten tutkija. Hänen alaansa ovat kielen variaatio ja muutosmekanismit, rakenteiden kompleksisuus ja typologia, Siperian kieliyhteisöt ja kielten dokumentaatio sekä sanojen alkuperätutkimuksen popularisoiminen kansanlaulun keinoin.
https://orcid.org/0000-0002-6400-7055

FT JUHA JANHUNEN on Helsingin yliopiston Itä-Aasian kielten ja kulttuurien professori. Hän on tehnyt kenttätyötä erityisesti Siperian, Mongolian ja Pohjois-Tiibetin alueen kielten parissa. Hän on tutkinut myös Koillis-Aasian kadonneita kirjakieliä, erityisesti kitaania.
https://orcid.org/0000-0003-1242-8808

FT LEENA KOLEHMAINEN on saksan kielen professori Turun yliopistossa. Hänen tutkimuksensa painopisteet liittyvät vertailevaan kielentutkimukseen, syntaksiin, kieli- ja käännöstieteen risteyskohtiin ja monikielisyyteen.
https://orcid.org/0000-0003-2746-7201

FT REETTA KONSTENIUS on tutkinut kielentutkimuksen metodologiaa suhteessa erilaisiin kognitiivisiin teorioihin Helsingin yliopistossa sekä Baskimaan ja Oviedon yliopistoissa Espanjassa. Hänen alaansa on erityisesti kielitieteen, teknologian ja digitaalisen tutkimuksen filosofia ja etiikka.

FT TOMMI KURKI on suomen kielen dosentti ja yliopistonlehtori Turun yliopistossa. Hän on sosiolingvisti, joka on tarkastellut puhutun kielen vaihtelua ja muutosta sekä alan metodologiaa eri tutkimuksissaan ja tutkimushankkeissa.
https://orcid.org/0000-0003-2858-9956

FT AKI-JUHANI KYRÖLÄINEN toimii tutkijana kielitieteen oppiaineessa McMasterin yliopistossa Kanadassa. Hänen tämänhetkinen tutkimuksensa tarkastelee elinikäisen oppimisen vaikutusta kieleen. Tutkimuksessaan Kyröläinen yhdistää sekä korpuslingvistisiä menetelmiä että koneoppimista kokeelliseen tutkimukseen.

FT VERONIKA LAIPPALA on digitaalisen kielentutkimuksen apulaisprofessori Turun yliopistossa. Hänen erityisalojaan ovat kielenkäyttö internetissä, koneoppimismenetelmien soveltaminen erityisen laajojen kieliaineistojen tutkimukseen sekä kieliaineistojen kehittäminen.
https://orcid.org/0000-0002-7635-429X

FT HANNA LAPPALAINEN on suomen kielen dosentti ja yliopistonlehtori Helsingin yliopistossa. Hän on kiinnostunut laajasti sosiolingvistiikasta, vuorovaikutuksesta sekä puhutteluun liittyvistä ilmiöistä.
https://orcid.org/0000-0001-5422-1667

FT HEINI LEHTONEN on pedagoginen yliopistonlehtori Helsingin yliopistossa. Hän on sosiolingvisti, jonka tutkimuskohteita ovat monikielinen vuorovaikutus, kielellinen moninaisuus ja kieli-ideologiat sekä monikielisyys koulutuksessa ja opetuksessa.
https://orcid.org/0000-0002-8726-0352

KTT ELLA LILLQVIST työskentelee tutkijatohtorina Helsingin yliopiston Kuluttajatutkimuskeskuksessa. Hän tutkii ideologiaa, eriarvoisuutta ja arjen selviytymisstrategioita myöhäiskapitalismin digitaalisissa konteksteissa soveltaen muun muassa kriittistä ja korpusavusteista diskurssianalyysia.
https://orcid.org/0000-0001-7419-4491

FT JOUKO LINDSTEDT on slaavilaisen filologian emeritusprofessori Helsingin yliopistossa. Hän on tutkinut Balkanin kieliliiton kieliä, erityisesti bulgariaa ja makedoniaa sekä muinaiskirkkoslaavia ja slaavien varhaishistoriaa, kielipolitiikkaa sekä esperanton äidinkielistymistä.
https://orcid.org/0000-0001-9021-7622

FT MILLA LUODONPÄÄ-MANNI on kielididaktiikan yliopistonlehtori Helsingin yliopistossa. Hänen erityisalaansa ovat kielentutkimuksen terminologia ja laajemmin lingvistiikan metodologia sekä monikielisyyteen liittyvät kysymykset.
https://orcid.org/0000-0002-1646-4462

FT MATTI MIESTAMO on yleisen kielitieteen professori Helsingin yliopistossa. Hänen alaansa ovat kielten laajamittaiseen vertailuun pohjautuva kielitypologia sekä vähän tutkittujen kielten dokumentaatio ja deskriptio. Hän on tutkinut erityisesti negaation ilmaisua maailman kielissä ja kielten kompleksisuuteen liittyviä kysymyksiä. Dokumentaatio- ja deskriptiotyötä hän on tehnyt koltansaamesta.
https://orcid.org/0000-0001-9801-9030

FT LIISA MUSTANOJA on suomen kielen yliopistonlehtori Tampereen yliopistossa. Hän on sosiolingvisti, jonka tutkimusta yhdistää kiinnostus yksilöihin ja heidän kielenkäyttöönsä muuttuvassa maailmassa. Mustanojan tutkimuskohteita ovat muun muassa Tampereen puhekieli, idiolektin muutos, reaaliaikatutkimuksen metodologia sekä uusimpana sota-ajan kirjeenvaihto.
https://orcid.org/0000-0003-1818-6926

FT JARKKO NIEMI toimii myynnin ja liiketoiminnan kehittämisen tutkijana Haaga-Helia ammattikorkeakoulussa. Myynnin ja asiakaspalvelun vuorovaikutuksen lisäksi hän on kiinnostunut arkikeskustelujen kieliopin tutkimuksesta ja keskustelunanalyysin yhdistämisestä kokeellisiin tutkimusasetelmiin.
https://orcid.org/0000-0001-6747-7821

FT TOMMI NIEMINEN on tutkinut suomen kielen fonetiikkaa, etenkin prosodiaa (nimenomaisesti puherytmiä ja intonaatiota). Hän on väitellyt Jyväskylän yliopistossa kielitieteen yleissemioottisesta tekstilajiteoriasta. Hänen muita viimeaikaisia kiinnostuksen kohteitaan ovat joukkoistettu lokalisointi sekä kirjoitusjärjestelmät.

FT URPO NIKANNE on suomen kielen professori Åbo Akademissa. Hänen alaansa on kieliopin ja kielellisen merkityksen teoria.
https://orcid.org/0000-0002-3267-6705

FT PIRKKO NUOLIJÄRVI on emeritaprofessori ja Kotimaisten kielten keskuksen aiempi johtaja. Hänen alaansa ovat sosiolingvistinen variaationtutkimus puhutussa kielessä ja kaunokirjallisuuden kielessä sekä kielipoliittisten ja kielten asemaan liittyvien kysymysten tarkastelu.
https://orcid.org/0000-0002-9688-6314

FT MICHAEL O'DELL on fonetiikan lehtori Tampereen yliopistossa. Hän on tarkastellut puheen rytmiä ja ajoitusta yleensä ja erityisesti rytmiä ja kvantiteettia suomen kielessä.
https://orcid.org/0000-0001-6496-2805

FT KRISTA OJUTKANGAS on suomen kielen dosentti ja yliopistonlehtori Turun yliopistossa. Hänen alaansa on kieliopin ja erityisesti spatiaalisen semantiikan tutkimus kognitiivisen kielentutkimuksen näkökulmasta.
https://orcid.org/0000-0002-4479-4637

FT MINNA PALANDER-COLLIN on englannin kielen professori Helsingin yliopistossa. Hänen alaansa ovat englannin kielen vaihtelu ja muutos, historiallinen sosiolingvistiikka ja pragmatiikka, kielellisen ja yhteiskunnallisen muutoksen suhteet sekä korpustutkimus.
https://orcid.org/0000-0001-8428-4423

FT GEDA PAULSEN on tutkija ja leksikografi Viron Eesti Keele Instituutissa. Hänen tutkimusaiheensa koskevat leksikaalista semantiikkaa, suomen kielen johto-oppia, suomen ja viron vertailua sekä soveltavaa kielitiedettä.
https://orcid.org/0000-0002-2749-4528

FT PEKKA PÄLLI on dosentti ja vanhempi yliopistonlehtori Aalto-yliopiston kauppakorkeakoulussa. Hän tutkii tekstejä ja kasvokkaista vuorovaikutusta osana organisaatioiden ja johtamisen käytäntöjä.
https://orcid.org/0000-0003-0287-8813

FT SARI PÖYHÖNEN on soveltavan kielentutkimuksen professori Jyväskylän yliopistossa. Hänen tutkimuksensa keskittyvät kielikoulutus-, maahanmuutto- ja kotouttamispolitiikkaan, kuulumisen narratiiveihin sekä kielen ja taiteen risteämiin.
https://orcid.org/0000-0003-1507-7206

FT JANNE SAARIKIVI on Helsingin yliopiston suomalais-ugrilaisten kielten ma. professori. Hän on tutkinut mm. suomalais-ugrilaista substraattia Pohjois-Venäjän murteissa, kielialueiden muuttumisen dynamiikkaa, vähemmistökielten säilymisen edellytyksiä sekä uralilaisten kielten sanaston historiaa (esim. suomen slaavilaislainoja ja permiläisten kielten itämerensuomalaisia lainoja).

FT ERIKA SANDMAN työskentelee yleisen kielitieteen yliopistonlehtorina Helsingin yliopistossa. Hän on kirjoittanut väitöskirjatyönään Länsi-Kiinassa puhutun, sinotiibetiläiseen kielikuntaan kuuluvan wutunin kielen referenssikieliopin. Hän on lisäksi tehnyt kenttätyötä Sahalinin saarella Venäjän Kaukoidässä uhanalaisen isolaattikieli nivhin puhujien parissa.
https://orcid.org/0000-0002-2341-2927

FT MINNA SEPPÄNEN on klassillisten kielten yliopistonlehtori Turun yliopistossa. Hän on tutkinut erityisesti kielen- ja kirjallisuudentutkimuksen kehitystä kreikkalais-roomalaisessa antiikissa sekä klassillisten kielten asemaa ja merkitystä autonomian ajan Suomessa.
https://orcid.org/0000-0001-6623-5708

FT KAIUS SINNEMÄKI on kvantitatiivisen ja vertailevan kielitieteen apulaisprofessori Helsingin yliopistossa. Hänen tutkimusalojaan ovat maailman kielten välinen vertailu eli kielitypologia, sosiolingvistinen typologia sekä kvantitatiiviset menetelmät kielentutkimuksessa.
https://orcid.org/0000-0002-6972-5216

FT PAULA SJÖBLOM on suomen kielen yliopistonlehtori ja dosentti Turun ylipistossa. Hänen alaansa ovat (kognitiivinen) nimistöntutkimus, tekstin- ja diskurssintutkimus sekä mainonnan ja talouselämän kieli.
https://orcid.org/0000-0002-7392-2300

FT JANNE SKAFFARI on englannin kielen yliopistonlehtori Turun yliopistossa. Hän opettaa kielihistoriaa, kielioppeja ja kieliasiantuntijuutta sekä tutkii erityisesti koodinvaihtoa ja lainasanoja muinais- ja keskienglannissa.
https://orcid.org/0000-0001-7349-9904

PHD MARJA-LEENA SORJONEN on suomen kielen professori Helsingin yliopistossa. Hänen keskeisiä kiinnostuksen kohteitaan ovat keskustelunanalyysi, vuorovaikutuslingvistiikka, variaationtutkimus, multimodaalisuus ja institutionaalinen vuorovaikutus.
https://orcid.org/0000-0002-8225-6805

FT TANJA SÄILY toimii englannin kielen apulaisprofessorina Helsingin yliopistossa. Hänen tutkimusalueisiinsa lukeutuvat korpuslingvistiikka, digitaaliset ihmistieteet, historiallinen sosiolingvistiikka, kielen vaihtelu ja muutos sekä englannin sanasto ja muoto-oppi.
https://orcid.org/0000-0003-4407-8929

FT LIISA TIITTULA on Helsingin yliopiston saksan kielen kääntämisen emeritaprofessori. Hänen tutkimusalueitaan ovat kirjallisuuden kääntäminen, kirjoitus- ja kuvailutulkkaus, multimodaalinen vuorovaikutus ja saksalais-suomalainen talousviestintä.
https://orcid.org/0000-0002-4688-0854

FT SEPPO VAINIO on erikoistutkija Turun yliopistossa. Hänen erityisalansa on psykolingvistiikka.

FT ULLA VANHATALO on yleisen kielitieteen dosentti Helsingin yliopistossa. Hänen alaansa ovat sanojen merkityksen tutkimus, yksinkertainen sanasto, merkityksen tutkimuksen menetelmät, selkokielen tutkimus ja kieliaineiden tutkijakoulutus.
https://orcid.org/0000-0002-2939-7836

FT TURO VARTIAINEN työskentelee tutkijatohtorina Helsingin yliopistossa. Hänen erikoisalojaan ovat englannin kielen vaihtelu ja muutos, historiallinen sosiolingvistiikka sekä konstruktiokielioppi.
https://orcid.org/0000-0002-4760-750X

VTT KIMMO VEHKALAHTI on soveltavan tilastotieteen dosentti ja Opettajien akatemian jäsen Helsingin yliopistossa. Hänen tutkimusalojaan ovat yhteiskuntadatatieteen soveltava opetus ja tutkimus, tilastolliset monimuuttujamenetelmät ja avoimen tieteen ohjelmistot.
https://orcid.org/0000-0001-6540-7328

FT MIKKO T. VIRTANEN on suomen kielen tutkijatohtori Helsingin yliopistossa. Hänen erikoisalanaan on tekstin- ja diskurssintutkimus. Hänen tutkimuksensa ovat käsitelleet muun muassa tarinankerrontaa tietokirjallisuudessa, mikro- ja makrotason toiminnan suhdetta tieteen teksteissä, topiikkien hallintaa monenkeskisessä pikaviestittelyssä sekä vanhuuden representaatioita uutismediassa.
https://orcid.org/0000-0002-8126-5929

FT MAX WAHLSTRÖM on eteläslaavilaisten kielten yliopistonlehtori Helsingin yliopistossa. Hän tutkii muun muassa argumentin merkintää ja viittaussuhteita kenttätyöaineistojen, historiallisten tekstien sekä typologisten piirreaineistojen avulla.
https://orcid.org/0000-0002-9585-5217

# Asiasanahakemisto

Hakemistossa on listattu kunkin sanan substantiivimuoto (esim. *realismi*), mutta viittaukset voivat sisältää myös hakusanan eri johdannaisia (kuten *realisti*, *realistisuus*, *realistinen*). Vakiintuneet termit, esimerkiksi *diskurssintutkimus*, ovat pääosin omina hakusanoinaan, mutta hakusanoja on voitu koota yläkäsitteiden alle, kuten sanan *tutkimus* alle. Rinnakkaistermit (esim. *epistemologia* ja *tieto-oppi*) on yleisesti merkitty omiksi hakusanoikseen kuitenkin ristiviittein.

**A**

**E**

**F**

**I**



**J**



**K**

**L**

**N**

**U**



**V**

**Y**



**Ä**

# Henkilöhakemisto

Henkilöhakemisto sisältää tekstissä mainitut viittaukset henkilöihin sekä lisäksi kirjoittajien ja toimittajien valikoimia viittauksia kielentutkimuksen eri alojen keskeisten julkaisujen kirjoittajiin.

**Z**

# Kielihakemisto

Kielihakemisto sisältää teoksessa viitattujen kielten lisäksi myös historialliset kielimuodot, rekonstruoidut kantakielet sekä viittaukset kielikuntiin ja kielialueisiin.